



Samuel Pacheco Marques Vaz Marcos

Licenciado em Matemática

Uma abordagem baseada em procura directa direccional a problemas com ruído

Dissertação para obtenção do Grau de Mestre em
**Matemática e Aplicações – Actuariado, Estatística e
Investigação Operacional**

Orientadora: Ana Luísa Custódio, Professora auxiliar,
Universidade NOVA de Lisboa

Júri

Presidente: Doutor Jorge Orestes Lasbarrères Cerdeira
Arguente: Doutora Maria do Carmo Proença Caseiro Brás
Vogal: Doutora Ana Luísa Custódio



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Novembro, 2016

Uma abordagem baseada em procura directa direcciona a problemas com ruído

Copyright © 2016 Samuel Pacheco Marques Vaz Marcos, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

Por decisão pessoal, o autor do texto não escreve segundo o novo acordo ortográfico.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

AGRADECIMENTOS

O mais sentido agradecimento à Professora Doutora Ana Luísa da Graça Batista Custódio pela orientação, acompanhamento, paciência e disponibilidade neste percurso de aprendizagem. Não há palavras que quantifiquem a dedicação e disponibilidade demonstrada.

Agradeço a todos os meus Professores das várias unidades curriculares do Mestrado.

A todos os meus colegas, principalmente ao Rui Cândido, que me acompanharam neste percurso académico um grande agradecimento pela amizade, motivação e companheirismo demonstrado.

À minha família, em especial à minha Mãe, um profundo agradecimento pela motivação, pela paciência e pela paz de espírito que criaram para que este trabalho fosse realizado.

RESUMO

Em aplicações a nível industrial, principalmente em engenharia, é frequente ser-se confrontado com problemas de optimização desafiadores onde, em particular, a avaliação da função objectivo pode estar contaminada com ruído numérico. A presença deste ruído inviabiliza a utilização de métodos de optimização baseados em derivadas.

A procura directa direccional (PDD) é um dos métodos que não recorre a derivadas. Nesta dissertação procurou-se avaliar o seu desempenho numérico na optimização de funções sujeitas a ruído, propondo eventuais adaptações que a tornassem mais eficiente/eficaz.

Foi feito um estudo numérico do seu desempenho considerando três tipos (baseado em polinómios de Chebyshev, Normal e Uniforme) e três níveis (5%, 10% e 20%) de ruído. Concluiu-se a necessidade da adaptação da PDD quando a inicialização considerada está longe do ponto óptimo, sobretudo na presença de níveis elevados de ruído, ou de ruído com oscilações mais irregulares (como é o caso do ruído Normal).

Adoptaram-se então abordagens não monótonas, onde se procura melhorar o valor da função objectivo não entre iterações consecutivas, mas ao longo de um histórico de iterações, permitindo escapar a mínimos locais espúrios, resultantes da presença do ruído.

Foram propostas duas variantes de PDD baseadas em abordagens não monótonas, analisadas as suas propriedades teóricas de convergência e o seu desempenho numérico. Estabeleceu-se a vantagem da utilização destes algoritmos, por comparação com algoritmos monótonos de PDD, para orçamentos computacionais moderados.

Analisou-se ainda as eventuais mais valias da implementação de uma ‘cache’. Em problemas onde o ruído é estocástico, esta mostrou-se dispensável. No caso de ruído baseado em polinómios de Chebyshev, dada a natureza determinística do ruído, o uso de uma ‘cache’ pode trazer algumas mais valias.

Palavras-chave: optimização sem recurso a derivadas, optimização sujeita a ruído, procura directa direccional, estratégias não monótonas

ABSTRACT

Commonly, in industrial applications, mainly in engineering, we face challenging optimization problems where, in particular, the evaluation of the objective function could be contaminated by numerical noise. The presence of this noise prevents the use of derivative based optimization methods.

Directional direct search (DDS) is one of the methods that does not resort to derivatives. In this thesis we tried to evaluate its numerical performance on the optimization of noisy functions, proposing adaptations that would make it more robust/efficient, if required.

A study on its numerical performance was conducted considering three types (based on Chebyshev polynomials, Normal, and Uniform) and three levels (5%, 10%, and 20%) of noise. We have concluded that DDS needs to be adapted when the initialization considered is far from the optimal point, mainly if the level of noise is high, or when in presence of noise with quite irregular oscillations (as it is the case of Normal noise).

Nonmonotone approaches were adopted, where the value of the objective function is not required to improve between consecutive iterations, but along an historic of iterations. This would allow to escape from spurious minima, resulting from the presence of noise.

Two variants of DDS based on non monotonous approaches were proposed, the corresponding theoretical properties related to convergence were analyzed, as well as their numerical performance. For moderate computational budgets, the advantage of using these algorithms was established, by comparison with monotonous variants of DDS.

The possible advantages of implementing a cache were also considered. In the presence of stochastic noise, the use of a cache is not required. When the noise results from Chebyshev polynomials, given its deterministic nature, the use of a cache could bring some improvement.

Keywords: derivative-free optimization, noisy optimization, directional direct search, nonmonotone strategies

ÍNDICE

Lista de Figuras	xiii
Lista de Tabelas	xv
1 Introdução - Optimização sem derivadas em problemas com ruído	1
1.1 Motivação	1
1.2 Trabalhos anteriores	2
1.3 Objectivos e organização da dissertação	4
2 A classe da procura directa direccional	5
2.1 Breve enquadramento histórico	5
2.2 Conjuntos geradores e bases positivas	5
2.3 Descrição de um algoritmo básico	6
2.4 Estratégias de globalização	9
2.5 Eficiência computacional	10
2.6 Análise de convergência	11
2.7 Tendências recentes da procura directa direccional	15
3 Análise de desempenho da procura directa direccional em optimização de problemas com ruído	17
3.1 Estudo exploratório do desempenho da procura directa direccional em quádras	17
3.2 Análise numérica do desempenho da procura directa direccional em problemas com ruído	23
3.2.1 A classe de problemas	24
3.2.2 Perfis de dados	26
3.2.3 Experiência numérica	26
4 Estratégias não monótonas	29
5 Estratégias não monótonas em procura directa direccional na optimização de problemas com ruído	31
5.1 Descrição das variantes algorítmicas	31
5.2 Análise de convergência	33

6	Análise de desempenho das estratégias não monótonas em PDD na optimização de problemas com ruído	37
6.1	Estratégias não monótonas	38
6.1.1	Ruído baseado em polinómios de Chebyshev	38
6.1.2	Ruído Normal	39
6.1.3	Ruído Uniforme	40
6.2	Impacto real do processo de optimização	41
6.3	Utilização de uma ‘cache’	42
7	Conclusões e questões em aberto	47
7.1	Conclusões	47
7.2	Questões em aberto	49
	Bibliografia	51
A	Tabelas	55
A.1	Estratégias não monótonas	55
A.1.1	Ruído baseado em polinómios de Chebyshev	56
A.1.2	Ruído Uniforme	57
A.1.3	Ruído Normal	59

LISTA DE FIGURAS

2.1	Um algoritmo de procura directa direccional.	9
2.2	Minimização de uma função recorrendo a um método de procura directa direccional.	11
3.1	Gráfico da função $f(x) = x^2 + 2x - 3$	18
3.2	Gráfico da função f (a vermelho) perturbada com um nível de ruído de 5%, baseado em polinómios de Chebyshev (a verde). Os asteriscos azuis representam os pontos obtidos pela PDD, correspondentes a iterações bem sucedidas.	19
3.3	Gráfico da função f (a vermelho) perturbada com um nível de ruído de 10%, baseado em polinómios de Chebyshev (a verde). Os asteriscos azuis representam os pontos obtidos pela PDD, correspondentes a iterações bem sucedidas.	20
3.4	Gráfico da função f (a vermelho) perturbada com um nível de ruído de 20%, baseado em polinómios de Chebyshev (a verde). Os asteriscos azuis representam os pontos obtidos pela PDD, correspondentes a iterações bem sucedidas.	20
3.5	Gráfico da função f (a vermelho) perturbada com um nível de ruído de 20%, baseado em ruído Normal (a verde). Os asteriscos azuis representam os pontos obtidos pela PDD, correspondentes a iterações bem sucedidas.	21
3.6	Gráfico da função f (a vermelho) perturbada com um nível de ruído de 20%, baseado em ruído Uniforme (a verde). Os asteriscos azuis representam os pontos obtidos pela PDD, correspondentes a iterações bem sucedidas.	21
3.7	Perfis de dados para o conjunto de problemas perturbados com ruído baseado em polinómios de Chebyshev, para os níveis de ruído de 5%, 10% e 20%, respectivamente.	27
3.8	Perfis de dados para o conjunto de problemas perturbados com ruído Uniforme, para os níveis de ruído de 5%, 10% e 20%, respectivamente.	28
3.9	Perfis de dados para o conjunto de problemas perturbados com ruído Normal para os níveis de ruído de 5%, 10% e 20%, respectivamente.	28
5.1	Construção e actualização do histórico de pontos nas estratégias não monótonas.	32
5.2	Um algoritmo de procura directa direccional, considerando estratégias não monótonas.	33
6.1	Perfis de dados para o conjunto de problemas perturbados com ruído baseado em polinómios de Chebyshev, para os níveis de ruído de 5%, 10% e 20%, com o comprimento do histórico de 5, 4 e 6 respectivamente.	38

6.2	Perfis de dados para o conjunto de problemas perturbados com ruído Normal, para os níveis de ruído de 5%, 10% e 20%, respectivamente, em qualquer dos casos com o comprimento do histórico igual a 2.	39
6.3	Perfis de dados para o conjunto de problemas perturbados com ruído Uniforme, para os níveis de ruído de 5%, 10% e 20%, respectivamente, em qualquer dos casos com o comprimento do histórico igual a 2.	41
6.4	Perfis de dados para o conjunto de problemas sem contaminação com ruído, mas utilizando as sequências de iteradas geradas anteriormente pelos algoritmos, para ruído baseado em polinômios de Chebyshev, ruído Normal e ruído Uniforme, respectivamente, considerando um nível de ruído de 10%.	42
6.5	Comparação do desempenho entre a PDD utilizando uma ‘cache’ e a PDD sem ‘cache’ considerando $\beta = 2$ para os dois primeiros perfis de dados e $\beta = 1$ para o terceiro.	43
6.6	Perfis de dados para o conjunto de problemas perturbados com os três tipos de ruído (baseado em polinômios de Chebyshev, Normal e Uniforme, respectivamente), considerando um nível de ruído de 10%, com $\beta = 2$	44
6.7	Perfis de dados para o conjunto de problemas perturbados com os três tipos de ruído (baseado em polinômios de Chebyshev, Normal e Uniforme, respectivamente), considerando um nível de ruído de 10%, com $\beta = 1$	44
6.8	Perfis de dados para o conjunto de problemas perturbados com os três tipos de ruído (baseado em polinômios de Chebyshev, Normal e Uniforme, respectivamente), considerando um nível de ruído de 10% e a implementação de uma ‘cache’.	45

LISTA DE TABELAS

3.1	Resultados da aplicação da PDD a problemas contaminados com ruído Uniforme, considerando três níveis de ruído.	22
3.2	Resultados da aplicação da PDD a problemas contaminados com ruído Normal, considerando três níveis de ruído.	22
3.3	Resultados da aplicação da PDD a problemas contaminados com ruído baseado em polinômios de Chebyshev, considerando três níveis de ruído.	22
3.4	Resultados da aplicação da PDD considerando um nível de ruído de 5% e os diferentes tipos de ruído.	23
3.5	Resultados da aplicação da PDD considerando um nível de ruído de 10% e os diferentes tipos de ruído.	23
3.6	Resultados da aplicação da PDD considerando um nível de ruído de 20% e os diferentes tipos de ruído.	23
3.7	Número de problemas para cada uma das dimensões, n_p , consideradas no conjunto teste ($2 \leq n_p \leq 12$).	24
3.8	Dimensões, valores ótimos encontrados na literatura e valores finais obtidos pela PDD considerando inicializações perto ou longe do ótimo.	25
A.1	Mínimo obtido para cada problema e para cada estratégia, considerando um nível de ruído de 5%.	56
A.2	Mínimo obtido para cada problema e para cada estratégia, considerando um nível de ruído de 10%.	56
A.3	Mínimo obtido para cada problema e para cada estratégia, considerando um nível de ruído de 20%.	57
A.4	Estatísticas descritivas da melhor estratégia (SUF_BWNMD), considerando um nível de ruído de 5%.	57
A.5	Estatísticas descritivas da melhor estratégia (SUF_BWNMD), considerando um nível de ruído de 10%.	58
A.6	Estatísticas descritivas da melhor estratégia (SUF_BWNMD), considerando um nível de ruído de 20%.	58
A.7	Estatísticas descritivas da melhor estratégia (SUF_BWNMD), considerando um nível de ruído de 5%.	59

A.8	Estatísticas descritivas da melhor estratégia (SUF_BWNMD), considerando um nível de ruído de 10%.	59
A.9	Estatísticas descritivas da melhor estratégia (SUF_NMD), considerando um nível de ruído de 20%.	60

INTRODUÇÃO - OPTIMIZAÇÃO SEM DERIVADAS EM PROBLEMAS COM RUÍDO

1.1 Motivação

Em aplicações industriais, nomeadamente em engenharia, é frequente a necessidade de resolver problemas de optimização, desafiantes a vários níveis. O aumento da capacidade computacional faz com que cada vez mais se recorra a modelos complexos para representação de sistemas, que frequentemente necessitam de ser optimizados [21, 26]. Neste contexto, não existe uma expressão analítica que represente a função objectivo e a sua avaliação numérica, com recurso ao modelo de simulação, é geralmente dispendiosa. A dedução analítica de expressões para as derivadas, essenciais à generalidade dos algoritmos de optimização numérica, é inviável e aproximações numéricas ao seu valor, recorrendo a diferenças finitas, são impraticáveis devido ao custo computacional associado à avaliação da função. Há casos ainda em que estas derivadas não existem nos pontos de interesse, devido ao carácter não suave da função.

Para além das dificuldades anteriormente descritas, a avaliação da função objectivo pode estar contaminada com ruído numérico, o que coloca desafios adicionais à optimização, já que a avaliação de um mesmo ponto pode conduzir a valores distintos fruto da contaminação com ruído. Aplicações de optimização em que a função objectivo se encontra contaminada com ruído surgem nas mais diversas áreas. Exemplos podem ser dados em aerodinâmica [21, 44], na área da medicina, mais concretamente no tratamento de tumores hepáticos e na simulação de modelos epidemiológicos para o cancro da mama [16], no transporte em gaseodutos [9] ou na definição de políticas económicas para recursos hídricos [31].

Métodos de optimização que não recorram a derivadas são então essenciais à resolução deste tipo de problemas, suscitando o interesse da comunidade científica e industrial pela

dificuldade inerente à classe de problemas a que se destinam e pelo elevado potencial de aplicação.

Estes algoritmos podem classificar-se em três grandes subclasses: os métodos de regiões de confiança baseados em interpolação ou regressão polinomial, os métodos de procura directa e os métodos de procura unidireccional [13]. No primeiro caso, em cada iteração do algoritmo a função objectivo é aproximada por um modelo de interpolação ou de regressão polinomial, válido numa vizinhança da iterada corrente (a região de confiança). Este modelo é optimizado, sendo a função objectivo avaliada no ponto resultante do processo de optimização. A eventual concordância entre o decréscimo sentido no modelo e a correspondente variação sentida na função objectivo dita a forma como deve ser actualizado o raio da vizinhança de validade do modelo e a própria actualização do modelo.

No caso dos métodos de procura directa distinguem-se a procura directa direccionada e a procura directa simplética. No primeiro caso, os algoritmos progridem por amostragem da função objectivo em conjuntos de pontos com uma geometria adequada. Esta classe de métodos será detalhada no Capítulo 2 do presente trabalho. Na procura directa simplética, de que é exemplo o famoso algoritmo do simplex de Nelder-Mead [38], a amostragem é feita em vértices de conjuntos simpléticos (conjuntos não degenerados de $n + 1$ pontos, num espaço de dimensão n). Em cada iteração, o algoritmo pretende melhorar o pior vértice, efectuando operações de reflexão, expansão e contracção do simplex que pretendem adaptar a sua forma à curvatura da função objectivo.

Por último, como a própria terminologia indica, os métodos de procura unidireccional consideram em cada iteração uma única direcção com boas propriedades de descida, que será explorada até que uma condição adequada de decréscimo do valor da função objectivo seja satisfeita.

Para uma abordagem mais detalhada de cada uma destas classes de métodos de optimização sem recurso a derivadas pode ser consultado o livro [13].

1.2 Trabalhos anteriores

No caso específico da optimização de funções sujeitas a ruído numérico, é comum considerar-se que os métodos de optimização sem recurso a derivadas permitem a obtenção de bons óptimos locais, para níveis moderados de ruído (ver Secção 1.3 de [13]).

Exemplos pertencentes à classe de métodos de regiões de confiança encontram-se nos trabalhos [8, 44]. Em [44], as considerações geométricas acerca da qualidade do conjunto de pontos usados na construção dos modelos são ignoradas na actualização do raio das regiões de confiança consideradas pelo algoritmo. O algoritmo é estendido à optimização sujeita a limites no valor das variáveis e aplicado num problema de optimização de forma em aerodinâmica.

Em [8] os autores consideram modelos de regressão com pesos que gerem o ruído presente na função objectivo. Uma outra abordagem baseada em regiões de confiança [34],

proposta pelos mesmos autores, consiste no ajuste dinâmico do comprimento do passo, dando passos maiores quando o modelo e a função estão de acordo entre si e passos mais pequenos quando o modelo é menos preciso. Este método não requer que o utilizador especifique um padrão fixo de pontos para construção dos modelos locais e não utiliza a reamostragem de pontos. Esta pode ser uma boa estratégia para lidar com ruído determinístico, onde o processo de reamostragem é totalmente ineficiente.

Contudo, quando o ruído tem uma natureza estocástica, a reamostragem de pontos foi utilizada em diversos trabalhos para adaptar os algoritmos à presença de ruído. Em [17] é desenvolvida uma variante do algoritmo UOBYQA (*Unconstrained Optimization BY Quadratic Approximation*) [40] para problemas de optimização na presença de ruído estocástico. Esta variante aplica técnicas Bayesianas para decidir se um dado passo do UOBYQA deve ou não ser executado. Estas alterações têm o intuito de controlar o erro aleatório através de médias de múltiplas avaliações por ponto, de forma a que o algoritmo proceda apropriadamente.

Analogamente, os mesmos autores apresentaram uma extensão do algoritmo DIRECT (*Dividing RECTangles*) [18] para resolver problemas de optimização global de funções sujeitas a ruído. O método DIRECT [28, 29] foi modificado, usando informação adquirida com amostras Bayesianas, para determinar um número apropriado de réplicas por ponto e através da sua média reduzir a incerteza presente na avaliação da função.

Métodos de procura unidireccional também foram sugeridos para otimizar problemas com ruído. Em [33] são utilizadas estratégias não monótonas, sendo escolhidas como direcções o gradiente espectral e as direcções de quasi-Newton, BFGS e SR1 [25]. Em qualquer dos casos, são feitas aproximações ao gradiente recorrendo a diferenças finitas centrais.

Num contexto em que a avaliação da função objectivo é dispendiosa, o uso de diferenças finitas pode ter um custo proibitivo. Uma solução de compromisso consiste na procura unidireccional usando gradientes simpléticos, também conhecida por filtro implícito [30]. De forma simplista, um gradiente simplético pode ser visto como o gradiente de um modelo linear de interpolação polinomial. A sua construção baseia-se então na amostragem da função objectivo em conjuntos simpléticos. Os conjuntos simpléticos estão também na base do famoso algoritmo de procura directa, o simplex de Nelder-Mead [38], para o qual existe uma variante adaptada a problemas sujeitos a ruído [2], novamente baseada em processos de reamostragem.

Refere-se ainda o uso de métodos de procura aleatória [10] que pretendem acelerar a optimização usando, para direcções de procura, vectores aleatórios. No caso do STARS (*Stepsize Approximation in Random Search*), o comprimento do passo é suavizado, de forma dinâmica e ajustada ao ruído presente na função objectivo, minimizando o erro dos mínimos quadrados entre a verdadeira derivada direcional da função com ruído e a sua aproximação com diferenças finitas.

Estratégias de suavização, não do comprimento do passo, mas da função objectivo, foram propostas em conjunto com a procura directa direcional, conduzindo a versões

robustas do algoritmo MADS (*Mesh Adaptive Direct Search*) [7]. Neste caso não é utilizada reamostragem de pontos, recorrendo-se a valores previamente avaliados para a função objectivo.

1.3 Objectivos e organização da dissertação

Da leitura da secção anterior conclui-se que são vários os trabalhos que analisam a utilização de métodos de optimização sem recurso a derivadas em problemas sujeitos a ruído. Contudo, pouco existe feito sobre procura directa direccional, uma classe de métodos que tem vindo a ser profundamente estudada, dadas as suas interessantes propriedades teóricas e os bons resultados em aplicações.

Neste trabalho pretende-se analisar o desempenho da procura directa direccional na optimização de funções sujeitas a ruído numérico, eventualmente equipando-a com uma abordagem não monótona que permita escapar aos óptimos locais espúrios resultantes da presença de ruído numérico.

No Capítulo 2 é feito um breve enquadramento histórico desta classe de métodos e a descrição de um algoritmo básico de procura directa direccional (PDD). É revista a análise de convergência destes algoritmos, detalhando-se os requisitos básicos para assegurar essa convergência, nomeadamente o uso de conjuntos geradores positivos e de estratégias de globalização. São ainda feitas considerações sobre a eficiência computacional destes algoritmos e tendências recentes desta classe de métodos.

A análise do desempenho numérico da PDD em problemas com ruído é feita no Capítulo 3, onde numa primeira abordagem se faz um estudo exploratório em quádras. Este estudo é posteriormente estendido a uma classe mais completa de problemas, que será utilizada nos restantes testes numéricos deste trabalho.

Uma resenha histórica do uso de estratégias não monótonas em optimização não linear é feita no Capítulo 4. Neste capítulo são descritas várias variantes não monótonas, bem como a sua evolução ao longo do tempo e a sua introdução na classe de algoritmos de optimização sem recurso a derivadas.

As estratégias não monótonas em procura directa direccional para optimização de funções sujeitas a ruído são apresentadas no Capítulo 5. Descrevem-se duas variantes algorítmicas de PDD não monótona e é feita a respectiva análise de convergência.

Estes algoritmos são numericamente testados no Capítulo 6. Neste capítulo é ainda feita uma experiência numérica, tentando melhorar a eficiência computacional da PDD não monótona, com a introdução de uma ‘cache’.

As principais conclusões deste trabalho são resumidas no Capítulo 7, onde ainda são mencionados alguns tópicos para investigação futura.

A CLASSE DA PROCURA DIRECTA DIRECCIONAL

2.1 Breve enquadramento histórico

O primeiro algoritmo de procura directa direccional conhecido é referido num relatório técnico de Fermi e Metropolis [20], em 1952. Contudo, o termo procura directa só foi proposto em 1961 por Hooke e Jeeves [27], e é apenas na década de 90, com a tese de doutoramento de Torczon [41] e os consequentes trabalhos [42, 43], que é feito um tratamento matemático para alguns algoritmos pertencentes a esta classe, nomeadamente para a procura em padrão. Esta classe de algoritmos é generalizada em 2002, por Audet e Dennis [4], surgindo pela primeira vez a organização de cada iteração num passo de procura e num passo de sondagem.

Os métodos de procura directa direccional caracterizam-se por em cada iteração não considerarem qualquer modelo explícito ou implícito da função objectivo, avaliando simplesmente o seu valor num conjunto de pontos com boas propriedades geométricas. Isto significa que, em cada iteração, o conjunto de pontos considerado deverá corresponder a pelo menos uma direcção de descida, independentemente da existência e/ou localização do vector gradiente. Os conjuntos geradores positivos ou as bases positivas são uma forma de definir estes conjuntos de pontos.

2.2 Conjuntos geradores e bases positivas

Definição 2.2.1. *Um conjunto de vectores $D = \{d_i, i = 1, \dots, r\} \subseteq \mathbb{R}^n$ diz-se um conjunto gerador positivo do conjunto $A \subseteq \mathbb{R}^n$ se qualquer vector de A pode ser escrito como combinação linear não negativa dos vectores d_i ($i = 1, \dots, r$), ou seja:*

$$A = \{d \in \mathbb{R}^n : d = \alpha_1 d_1 + \alpha_2 d_2 + \dots + \alpha_r d_r, \alpha_i \in \mathbb{R}_0^+\}.$$

Uma base positiva em \mathbb{R}^n é um conjunto gerador positivo positivamente independente, ou seja, é um conjunto de vectores que gera \mathbb{R}^n através de combinações lineares não negativas e tal que nenhum dos seus vectores pode ser obtido como combinação linear não negativa dos restantes. A cardinalidade de uma base positiva de \mathbb{R}^n está compreendida entre $n + 1$ e $2n$ elementos [15]. Uma base positiva com cardinalidade $n + 1$ designa-se por base positiva minimal, enquanto que uma base positiva que possua cardinalidade $2n$ diz-se uma base positiva maximal.

Exemplos de bases positivas minimais e maximais de \mathbb{R}^n , são respectivamente:

$$\begin{bmatrix} 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & -1 \\ 0 & 0 & \dots & 1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

Num contexto de optimização onde não é possível utilizar derivadas, mesmo que elas existam, qualquer conjunto gerador positivo contém pelo menos uma direcção com boas propriedades de descida. De facto, assumindo que ∇f existe, mas não está disponível para uso, prova-se que pelo menos uma das direcções de um conjunto gerador positivo faz um ângulo inferior a 90° com $-\nabla f$, ou seja, é uma direcção de descida [15].

Proposição 2.2.1. *O conjunto $\{d_1, \dots, d_r\}$ gera positivamente \mathbb{R}^n se e só se para qualquer vector $v \neq 0$ em \mathbb{R}^n é possível encontrar $i \in \{1, \dots, r\}$ tal que $v^\top d_i > 0$.*

Esta proposição garante que, se se tiver uma função diferenciável numa vizinhança de um ponto x para o qual $\nabla f(x) \neq 0$, ao explorar os vectores pertencentes a um conjunto gerador positivo de \mathbb{R}^n , um desses vectores d_i satisfaz $-\nabla f(x)^\top d_i > 0$. Isto significa que, quando $\nabla f(x) \neq 0$, todo o conjunto gerador positivo de \mathbb{R}^n apresenta pelo menos uma direcção de descida.

2.3 Descrição de um algoritmo básico

A descrição adoptada para a classe de métodos de procura directa direcciona segue a estrutura proposta por Audet e Dennis [4], no contexto da procura em padrão generalizada.

Considere-se um problema de optimização sem restrições:

$$\min f(x), \text{ com } f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\},$$

onde se assume que as derivadas de f ou não existem, ou não estão disponíveis para uso, não podendo ser aproximadas numericamente. Um algoritmo de procura directa direccional gera uma sucessão de iteradas $\{x_k\}_{k \in K} \subseteq \mathbb{R}^n$, cuja correspondente sucessão de valores para a função objectivo $\{f(x_k)\}_{k \in K}$ é decrescente em sentido lato. Cada iteração divide-se em dois passos: o passo de procura e o passo de sondagem.

Passo de procura

O passo de procura é opcional e desnecessário para as propriedades de convergência do método. Este passo é normalmente definido pelo utilizador, frequentemente correspondendo a uma heurística que já era usada. Como requisito geral, em cada iteração terá que avaliar a função objectivo num número finito de pontos. Contudo, dependendo da estratégia de globalização utilizada, que será detalhada na Secção 2.4, os pontos avaliados poderão ter de pertencer a uma grelha implícita considerada pelo algoritmo. O passo de procura pode ainda ser usado para a implementação de heurísticas que permitem melhorar a eficiência algorítmica. Este passo diz-se bem sucedido quando se encontra um ponto que corresponde a um melhor valor da função objectivo. Se tal não ocorrer, o algoritmo obrigatoriamente executa o passo de sondagem.

Passo de sondagem

Neste segundo passo, a função objectivo é avaliada numa vizinhança discreta da iterada corrente (x_k), definida à custa de uma base positiva ou de um conjunto gerador positivo – o chamado *conjunto de sondagem*.

Definição 2.3.1. *Seja D um conjunto finito de bases positivas ou de conjuntos geradores positivos de \mathbb{R}^n . O conjunto de sondagem P_k , associado à iterada x_k de um método de procura directa direccional é definido por:*

$$P_k = \{x_k + \Delta_k d : d \in D_k\},$$

onde Δ_k representa um comprimento do passo e $D_k \in D$. O vector $d \in D_k$ diz-se uma direcção de sondagem e $x_k + \Delta_k d \in P_k$ um ponto de sondagem.

Se existir um ponto de sondagem onde o valor da função objectivo é melhor do que o da iterada corrente ($f(x_k + \Delta_k d) < f(x_k)$) então o passo de sondagem diz-se bem sucedido.

Caso contrário, estar-se-á na presença de um insucesso.

No teste do conjunto de sondagem pode-se seguir uma estratégia oportunista ou completa. No primeiro caso aceita-se um ponto de sondagem logo que se encontre um ponto que reduz o valor da função objectivo, não avaliando os pontos correspondentes às restantes direcções de sondagem. No caso da estratégia completa, todas as direcções de sondagem são testadas, escolhendo-se como nova iterada o ponto correspondente ao valor mais baixo que melhore a função objectivo (caso exista). Note-se que no caso de um passo de sondagem mal sucedido, as duas estratégias são indistintas. Uma iteração dir-se-à bem sucedida se o passo de procura ou o passo de sondagem forem bem sucedidos. Caso contrário, dir-se-à um insucesso.

Actualização do comprimento do passo

Dependendo da existência de sucesso ou insucesso, o comprimento de passo, Δ_k , deverá ser actualizado no final de cada iteração.

Para tal, fixe-se inicialmente $\beta \geq 1$ e $0 < \mu < 1$. Quando a iteração é bem sucedida, o comprimento do passo é mantido ou aumentado ($\Delta_{k+1} = \beta \Delta_k$). Na presença de um insucesso, o comprimento de passo é obrigatoriamente reduzido ($\Delta_{k+1} = \mu \Delta_k$).

Uma descrição algorítmica da procura directa direccional encontra-se na Figura 2.1.

Inicialização

Considerar $x_0 \in \mathbb{R}^n$ com $f(x_0) < +\infty$, um conjunto de conjuntos geradores positivos ou de bases positivas D de \mathbb{R}^n e um comprimento inicial de passo $\Delta_0 > 0$. Considerar ainda $\beta \geq 1$ e $0 < \mu < 1$, utilizados na actualização do comprimento do passo. Seja $k = 0$.

Passo de Procura

Avaliar a função objectivo num número finito de pontos, tentando encontrar um ponto $x \in \mathbb{R}^n$ que satisfaça $f(x) < f(x_k)$. Se tal for possível, definir a iteração como bem sucedida, considerar $x_{k+1} = x$ e omitir o passo de sondagem. Caso contrário, realizar o passo de sondagem.

Passo de Sondagem

Escolher $D_k \in D$ e considerar o conjunto de sondagem $P_k = \{x_k + \Delta_k d : d \in D_k\}$. Avaliar f em P_k .

Se existir $x_k + \Delta_k d \in P_k$ tal que $f(x_k + \Delta_k d) < f(x_k)$, considerar $x_{k+1} = x_k + \Delta_k d$ e definir a iteração como bem sucedida. Caso contrário, considerar $x_{k+1} = x_k$ e definir a iteração como um insucesso.

Actualização do Comprimento do Passo

Se a iteração foi bem sucedida, fazer $\Delta_{k+1} = \beta \Delta_k$. Caso contrário, considerar $\Delta_{k+1} = \mu \Delta_k$. Incrementar o contador de iterações, fazendo $k = k + 1$ e voltar ao passo de procura.

Figura 2.1: Um algoritmo de procura directa direccional.

2.4 Estratégias de globalização

Para garantir algum tipo de convergência para a sucessão de iteradas gerada por um algoritmo de procura directa direccional é necessário mostrar que existe pelo menos uma subsucessão de iteradas convergente. Para tal, um primeiro passo consiste em garantir que uma subsucessão de comprimentos de passo converge para zero. Esta propriedade é obtida com as chamadas estratégias de globalização, que podem ser de dois tipos: uso de grelhas racionais ou imposição de decréscimo suficiente.

No primeiro caso, quando a estratégia de globalização é baseada na utilização de grelhas racionais, todos os pontos gerados pelo algoritmo devem pertencer a uma grelha implícita

$$M_k = \{x_k + \Delta_k D z : z \in \mathbb{N}_0^{|D|}\},$$

em que D é uma matriz cujas colunas correspondem aos vectores de um número finito de conjuntos geradores positivos, $|D|$ representa a sua cardinalidade e \mathbb{N}_0 é o conjunto de inteiros maiores ou iguais a zero. Para garantir esta propriedade, há que impor algumas condições na forma como são construídas as bases positivas ou os conjuntos geradores positivos usados nas várias iterações do algoritmo e também na forma como são feitas as actualizações do comprimento do passo [4]. Esta estratégia permite a aceitação de novos pontos desde que haja um decréscimo simples no valor da função objectivo.

Alternativamente, a estratégia de globalização pode ser baseada na imposição de uma condição de decréscimo suficiente. Neste caso, a flexibilidade presente nas regras de actualização do comprimento de passo e na construção das direcções de sondagem tem como contrapartida uma maior exigência no critério de aceitação de novos pontos. Nesta variante, uma iterada é bem sucedida se a condição

$$f(x_{k+1}) < f(x_k) - \rho(\Delta_k) \quad (2.1)$$

for satisfeita, onde ρ representa uma função definida em $]0, +\infty[$, contínua, positiva e crescente, satisfazendo

$$\lim_{t \rightarrow 0^+} \frac{\rho(t)}{t} = 0.$$

Um exemplo de uma família de funções com estas características é a definida por $\rho(\alpha) = c\alpha^j$, com $c > 0$, $j > 1$ constantes reais [32].

2.5 Eficiência computacional

A grande desvantagem do uso da procura directa direcciona prende-se com a sua eficiência numérica. Os algoritmos são rápidos a decrescer o valor inicial, mas a convergência é lenta quando o algoritmo está próximo do ponto óptimo, como pode ser observado na Figura 2.2.

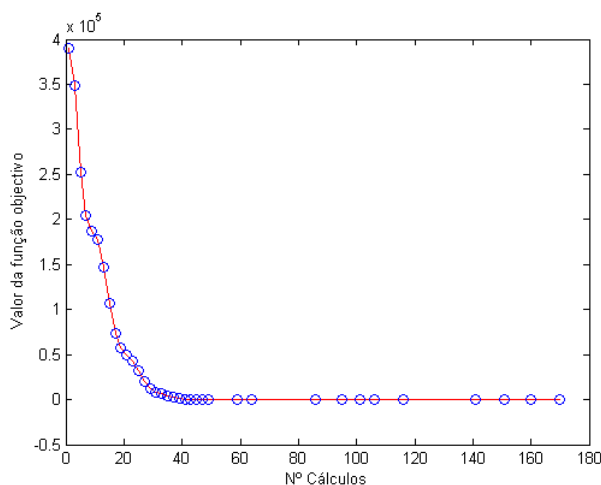


Figura 2.2: Minimização de uma função recorrendo a um método de procura directa direccional.

Para aumentar esta eficiência, uma estratégia simples consiste na implementação de uma ‘cache’, ou seja, de um histórico de pontos avaliados, por forma a impedir que o algoritmo volte a testar pontos cujo valor da função objectivo já tenha sido calculado. Antes da nova avaliação, o algoritmo verifica se o ponto em causa não consta já da ‘cache’, evitando qualquer avaliação desnecessária da função objectivo.

Outras estratégias contemplam a ordenação das direcções de sondagem, testando, por exemplo, em primeiro lugar a última direcção bem sucedida. Versões mais elaboradas consistem na definição de indicadores de descida, sendo as direcções de sondagem ordenadas de acordo com o menor ângulo que fazem com estes indicadores [14].

A utilização de vectores de comprimento de passo em vez de um parâmetro escalar [6] pode também ajudar a reduzir a sensibilidade da procura directa direccional ao escalonamento das variáveis, contribuindo para o aumento da eficiência computacional.

2.6 Análise de convergência

Para esta classe de métodos de optimização sem recurso a derivadas pode ser deduzida uma hierarquia de resultados de convergência, que dependem da suavidade da função objectivo [4]. Mesmo no caso de funções descontínuas, pode-se estabelecer garantias de alguma forma de estacionaridade [45].

A análise de convergência apresentada baseia-se no trabalho de Audet e Dennis [4] e de Kolda, Lewis e Torczon [32]. A primeira hipótese considerada por estes autores pressupõe que todas as iteradas produzidas por um algoritmo de procura directa direccional pertencem a um conjunto compacto. Para tal, uma condição suficiente seria supor que o conjunto

$$L(x_0) = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\} \quad (2.2)$$

é compacto. No entanto, em alguns resultados basta assumir que $L(x_0)$ é limitado.

Hipótese 2.6.1. *O conjunto $L(x_0) = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ é limitado.*

Em primeiro lugar, terá que se provar a existência de uma subsucessão de comprimentos de passo que converge para zero, garantindo assim, a convergência de subsucessões de pontos gerados pelo algoritmo.

Teorema 2.6.1. *Sob a Hipótese 2.6.1, um método de procura directa direcciona, considerando uma estratégia de globalização baseada em grelhas racionais, gera uma sucessão de comprimentos de passo tal que,*

$$\liminf_{k \rightarrow +\infty} \Delta_k = 0.$$

Demonstração. Suponha-se que

$$\liminf_{k \rightarrow +\infty} \Delta_k \neq 0.$$

Então existe um $\Delta^* > 0$, tal que $\Delta_k > \Delta^*, \forall k \in \mathbb{N}$. Todos os pontos gerados pelo algoritmo pertencem a uma grelha racional implícita que constitui um reticulado [4]. A intersecção de um conjunto limitado com um reticulado é finita. Assim, há um número finito de pontos distintos que se podem aceitar como novas iteradas em iterações bem sucedidas, o que significa que há um número finito de sucessos e infinitos insucessos. Em cada iteração mal sucedida o comprimento de passo é reduzido em $\mu > 0$, o que conduz a um absurdo.

□

No caso da estratégia de globalização do algoritmo ser baseada numa condição de decréscimo suficiente, o conjunto das imagens é que terá que ser limitado.

Hipótese 2.6.2. *O conjunto $\{f(x) : f(x) \leq f(x_0)\}$ é limitado inferiormente.*

Teorema 2.6.2. *Sob a Hipótese 2.6.2, um método de procura directa direcciona, que usa a condição (2.1) para exigir decréscimo suficiente nas iterações com sucesso, gera uma sucessão de comprimentos de passo tal que,*

$$\liminf_{k \rightarrow +\infty} \Delta_k = 0.$$

Demonstração. Suponha-se que

$$\liminf_{k \rightarrow +\infty} \Delta_k \neq 0.$$

Então existe um $\Delta^* > 0$, tal que $\Delta_k > \Delta^*, \forall k \in \mathbb{N}$. Dado que em cada iteração sem sucesso o factor utilizado para a actualização comprimento de passo é $0 < \mu < 1$, o número de iterações com sucesso é infinito. Atendendo às propriedades da função $\rho(\cdot)$, tem-se $\rho(\Delta_k) \geq \rho(\Delta^*) > 0$. Pode-se assim afirmar que $f(x_{k+1}) - f(x_k) < -\rho(\Delta_k) \leq -\rho(\Delta^*) < 0$ para iterações bem sucedidas. Para iterações sem sucesso, $f(x_{k+1}) - f(x_k) = 0$, o que leva a concluir que $f(x_k) \rightarrow -\infty$ contradizendo a Hipótese 2.6.2.

□

Dado que

$$\liminf_{k \rightarrow +\infty} \Delta_k = 0$$

e que o comprimento de passo só é reduzido em insucessos tem que existir uma subsucessão $\{x_k\}_{k \in K'}$, correspondente a iterações sem sucesso, tal que

$$\lim_{k \in K'} \Delta_k = 0.$$

Uma subsucessão com estas características designa-se por *subsucessão refinada*. Como $L(x_0)$ pela Hipótese 2.6.1 é limitado, existe uma subsucessão de $\{x_k\}_{k \in K'}$ que adicionalmente converge. Ou seja, existe uma subsucessão de iteradas sem sucesso $\{x_k\}_{k \in K''}$ e $K'' \subseteq K'$ tal que

$$\lim_{k \in K''} \Delta_k = 0 \quad e \quad \lim_{k \in K''} x_k = x^*.$$

A convergência da procura directa direccionada num contexto não suave é estabelecida ao analisar o comportamento do algoritmo em pontos limite de subsucessões refinadas. Para tal, começa-se por introduzir o conceito de derivada direccionada generalizada de Clarke [12], definida para funções localmente contínuas à Lipschitz, ou seja, que localmente satisfazem uma condição do tipo $|f(x) - f(y)| \leq L\|x - y\|$, com $L \geq 0$.

Definição 2.6.1. *Seja f uma função continua à Lipschitz perto de x^* . A derivada direccionada generalizada de Clarke da função f , calculada em x^* , na direcção d é dada por*

$$f^o(x^*; d) = \limsup_{\substack{x \rightarrow x^* \\ t \downarrow 0}} \frac{f(x + td) - f(x)}{t}.$$

Recorrendo à derivada generalizada de Clarke, estabelece-se o primeiro resultado de convergência para a procura directa direccionada.

Teorema 2.6.3. *Seja x^* o ponto limite de uma subsucessão refinada convergente gerada com um método de procura directa direcciona e f continua à Lipschitz numa vizinhança de x^* . Suponha-se ainda que f é avaliada num número infinito de vezes na direcção d usada na definição dos conjuntos de sondagem da subsucessão refinada. Então a derivada direcciona generalizada de Clarke da função f definida em x^* , na direcção d , é não negativa, ou seja,*

$$f^o(x^*; d) \geq 0.$$

Demonstração. Seja $\{x_k\}_{k \in K}$ uma subsucessão refinada convergente para x^* . Dado que f é contínua à Lipschitz numa vizinhança de x^* , quando se considera uma estratégia de globalização baseada em grelhas racionais, pela definição de Clarke tem-se,

$$f^o(x^*; d) = \limsup_{\substack{x \rightarrow x^* \\ t \downarrow 0}} \frac{f(x + td) - f(x)}{t} \geq \lim_{k \in K} \frac{f(x_k + \Delta_k d) - f(x_k)}{\Delta_k}.$$

Este último quociente é não negativo, dado que x_k é um elemento da subsucessão refinada convergente para x^* , logo correspondente a uma iteração sem sucesso. Assim sendo, o limite superior inicial é também não negativo.

Se a estratégia de globalização for baseada numa condição de decréscimo suficiente, dado que $\frac{\rho(\Delta_k)}{\Delta_k} \rightarrow 0$, quando $\Delta_k \rightarrow 0$, tem-se

$$\begin{aligned} f^o(x^*; d) &= \limsup_{\substack{x \rightarrow x^* \\ t \downarrow 0}} \frac{f(x + td) - f(x)}{t} \geq \\ \lim_{k \in K} \frac{f(x_k + \Delta_k d) - f(x_k)}{\Delta_k} &= \lim_{k \in K} \frac{f(x_k + \Delta_k d) - f(x_k) + \rho(\Delta_k)}{\Delta_k} - \frac{\rho(\Delta_k)}{\Delta_k} \geq 0. \end{aligned}$$

□

Uma versão mais robusta do resultado anterior pode ser obtida se se assumir mais suavidade acerca da função objectivo. Note-se que, como D tem cardinalidade finita, haverá sempre um conjunto gerador positivo que será utilizado um número infinito de vezes em cada subsucessão refinada.

Teorema 2.6.4. *Seja x^* o ponto limite de uma subsucessão refinada convergente gerada com um método de procura directa direcciona. Se f for estritamente diferenciável em x^* , então*

$$\nabla f(x^*) = 0.$$

Demonstração. Se f é diferenciável em x^* , então $f^o(x^*; v) = \nabla f(x^*)^\top v$ [12]. Seja D' o conjunto gerador positivo que é usado um número infinito de vezes pela subsucessão refinada. Pelo Teorema 2.6.3, para cada $d \in D'$, $\nabla f(x^*)^\top d \geq 0$. Considere-se $v \in \mathbb{R}^n$, não nulo e não ortogonal a $\nabla f(x^*)$. O vector v pode ser escrito como combinação linear não negativa dos elementos de D' , ou seja, $v = \sum_{i=1}^{|D'|} \alpha_i d_i$ com $\alpha_i \geq 0$. Assim $\nabla f(x^*)^\top v = \sum_{i=1}^{|D'|} \alpha_i \nabla f(x^*)^\top d_i \geq 0$. Considerando agora $-v$, repetindo o raciocínio anterior, obtém-se $-\nabla f(x^*)^\top v \geq 0$, logo $\nabla f(x^*)^\top v = 0$. Sendo v não nulo e não ortogonal a $\nabla f(x^*)$ vem $\nabla f(x^*) = 0$. □

2.7 Tendências recentes da procura directa direccional

Quando se pretende a minimização de uma função não suave recorrendo a um método de procura directa direccional, o uso de um número finito de conjuntos geradores ou de bases positivas pode conduzir a situações de convergência para pontos que não correspondem ao mínimo da função. Alguns exemplos foram apresentados em [32], motivando o desenvolvimento de uma nova classe de algoritmos de procura directa direccional – *MADS (Mesh Adaptive Direct Search)* [5].

Continuando a usar uma estratégia de globalização baseada em grelhas racionais, a característica distintiva desta nova classe de algoritmos resulta da reunião dos conjuntos geradores ou das bases positivas usadas ao longo das várias iterações ser assintoticamente densa na esfera unitária. A implementação original [5] de um conjunto de direcções de sondagem com as características mencionadas baseava-se em matrizes triangulares inferiores (*LTMADS*) e envolvia alguma aleatoriedade. Posteriormente, por forma a evitar esta aleatoriedade e a obter uma melhor distribuição espacial das direcções que constituem cada conjunto gerador positivo, foi proposta a implementação *OrthoMADS* [1], onde as direcções que constituem cada conjunto de sondagem são ortogonais entre si e geradas de forma determinística.

Recentemente, com o objectivo de aumentar a eficiência computacional da procura directa direccional, foi proposta a procura directa direccional probabilística [22]. Neste caso, em cada iteração são considerados conjuntos de sondagem que não são necessariamente bases ou conjuntos geradores positivos. Apenas têm de garantir descida probabilística, ou seja, que com uma certa probabilidade pelo menos um desses vectores faz um ângulo agudo com o simétrico do vector gradiente. O processo de sondagem pode assim utilizar menos do que $n + 1$ direcções (cardinalidade mínima de uma base positiva), permitindo ganhos na eficiência computacional. A convergência é neste caso garantida em probabilidade.

ANÁLISE DE DESEMPENHO DA PROCURA DIRECTA DIRECCIONAL EM OPTIMIZAÇÃO DE PROBLEMAS COM RUÍDO

Neste capítulo pretende-se estabelecer a eventual necessidade da definição de estratégias de adaptação da procura directa direccional (PDD) à optimização de problemas com ruído. Para tal, começou-se por analisar o seu desempenho numérico na presença de diferentes tipos e níveis de ruído, procurando-se ainda estudar a dependência da inicialização considerada. Em qualquer dos testes numéricos realizados o algoritmo de PDD descrito na Figura 2.1 foi implementado em Matlab, considerando-se o conjunto gerador positivo $[e_n, -e_n, I_n, -I_n]$, em que e_n é um vector coluna de uns de dimensão n e I_n representa a matriz identidade. Considerou-se uma estratégia de globalização baseada em grelhas racionais, um comprimento de passo inicial $\Delta_0 = 1$ e os parâmetros $\beta = 2$ e $\mu = \frac{1}{2}$ na actualização do comprimento de passo. O passo de sondagem foi efectuado de forma oportunista, tal como descrito na Secção 2.3. Como critério de paragem permitiu-se um máximo de 1500 avaliações da função.

3.1 Estudo exploratório do desempenho da procura directa direccional em quádricas

Numa primeira abordagem procurou-se ganhar sensibilidade ao impacto que diferentes tipos de ruído têm na função objectivo e à capacidade da PDD em determinar a solução dos problemas resultantes da contaminação com estes diferentes tipos de ruído. Assim, começou-se por considerar a função quadrática $f(x) = x^2 + 2x - 3$. Optou-se por uma quádrica de dimensão um, que facilmente permite uma representação gráfica (ver

Figura 3.1), sendo uma função simples, cujo mínimo global é conhecido (neste caso corresponde ao par $(-1, -4)$).

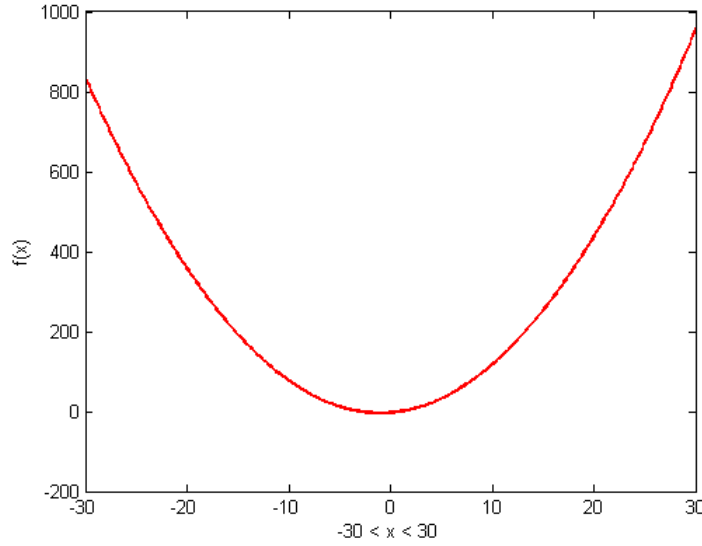


Figura 3.1: Gráfico da função $f(x) = x^2 + 2x - 3$.

A quádrlica anterior foi contaminada com três tipos distintos de ruído: baseado em polinómios de Chebychev [36], Uniforme [36] ou Normal [33]. Para cada tipo de problemas, definido pelo ruído considerado, analisaram-se ainda três níveis de ruído distintos: 5%, 10% e 20%.

A classe de problemas com ruído Uniforme [36] é dada por:

$$\bar{f}(x) = (1 + U)f(x), \quad (3.1)$$

em que U é uma variável com distribuição uniforme no intervalo $[-\varepsilon, \varepsilon]$, sendo ε o nível de ruído considerado.

A classe de problemas com ruído baseado em polinómios de Chebyshev [36] é definida por:

$$\bar{f}(x) = (1 + \varepsilon\alpha(x))f(x), \quad (3.2)$$

em que novamente ε representa a percentagem de ruído considerada,

$$\alpha(x) = \alpha_0(x)(4\alpha_0(x)^2 - 3) \quad (3.3)$$

e

$$\alpha_0(x) = 0,9 \sin(100\|x\|_1) \cos(100\|x\|_\infty) + 0,1 \cos(\|x\|_2). \quad (3.4)$$

3.1. ESTUDO EXPLORATÓRIO DO DESEMPENHO DA PROCURA DIRECTA DIRECCIONAL EM QUÁDRICAS

Finalmente, a classe de problemas com ruído Normal [33] é definida por:

$$\tilde{f}(x) = (1 + N)f(x), \quad (3.5)$$

com N uma variável com distribuição Normal de média $\mu = 0$ e variância $\sigma^2 = \varepsilon^2$, sendo ε a percentagem de ruído considerada.

Adicionalmente ao máximo de 1500 cálculos para a função objectivo, considerou-se um comprimento mínimo de passo igual a $2.220E - 16$ (correspondente à precisão de máquina utilizada pelo Matlab).

Os gráficos das Figuras 3.2 e 3.3 ilustram o impacto do aumento de 5% na percentagem de ruído considerado, quando este é baseado em polinómios de Chebyshev.

A análise dos dois gráficos seguintes torna claro o efeito da estrutura do ruído considerada. À medida que há um distanciamento do ponto óptimo, verifica-se que existe uma maior oscilação na avaliação da função objectivo. Observa-se também que um aumento de apenas 5% no nível de ruído (em relação à Figura 3.2) tem um impacto significativo no nível de oscilação da função. Ainda assim, a PDD é bem sucedida na determinação do ponto óptimo.

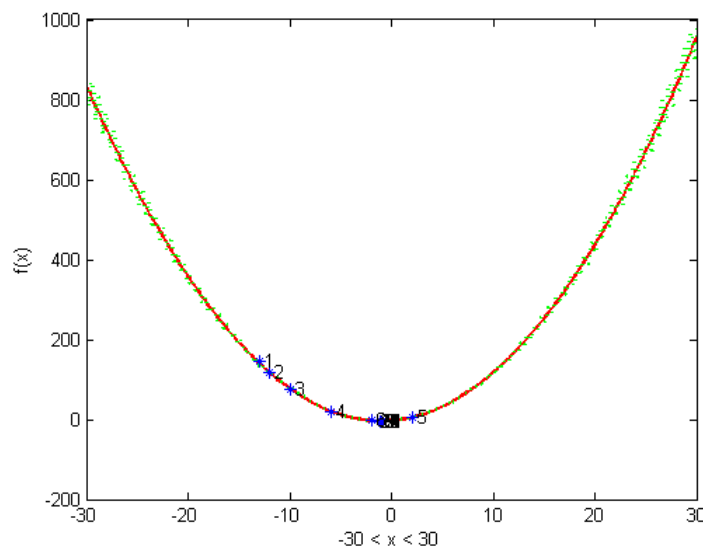


Figura 3.2: Gráfico da função f (a vermelho) perturbada com um nível de ruído de 5%, baseado em polinómios de Chebyshev (a verde). Os asteriscos azuis representam os pontos obtidos pela PDD, correspondentes a iterações bem sucedidas.

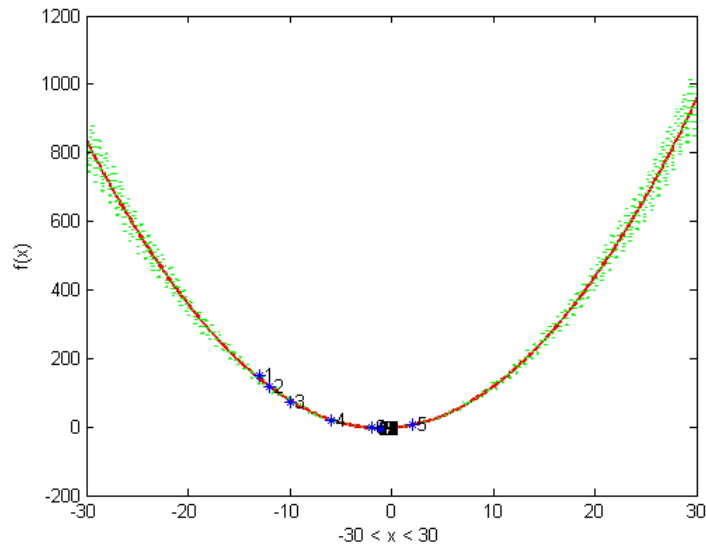


Figura 3.3: Gráfico da função f (a vermelho) perturbada com um nível de ruído de 10%, baseado em polinômios de Chebyshev (a verde). Os asteriscos azuis representam os pontos obtidos pela PDD, correspondentes a iterações bem sucedidas.

Considerando agora o nível de ruído mais elevado (20%), as Figuras 3.4– 3.6 procuram ilustrar cada um dos três tipos de ruído adoptados (baseado em polinômios de Chebyshev, ruído Uniforme e ruído Normal, respectivamente).

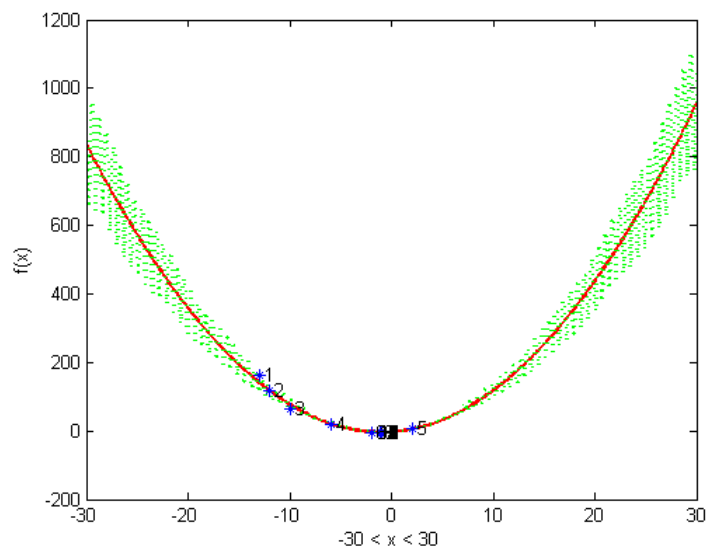


Figura 3.4: Gráfico da função f (a vermelho) perturbada com um nível de ruído de 20%, baseado em polinômios de Chebyshev (a verde). Os asteriscos azuis representam os pontos obtidos pela PDD, correspondentes a iterações bem sucedidas.

3.1. ESTUDO EXPLORATÓRIO DO DESEMPENHO DA PROCURA DIRECTA DIRECCIONAL EM QUÁDRICAS

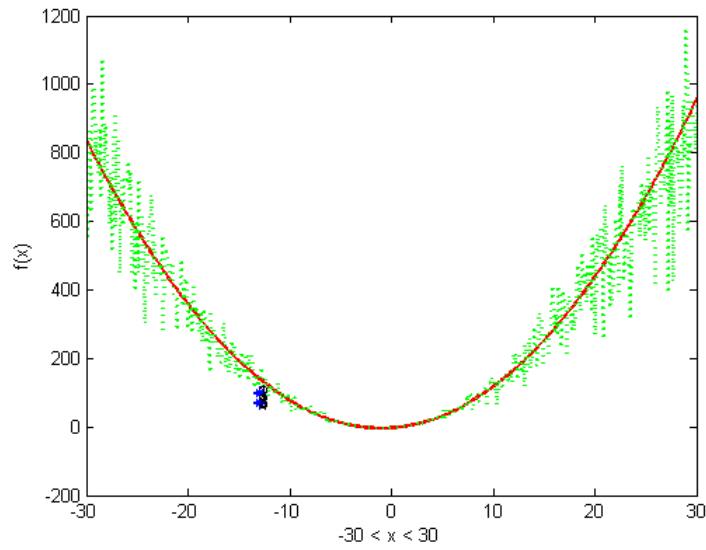


Figura 3.5: Gráfico da função f (a vermelho) perturbada com um nível de ruído de 20%, baseado em ruído Normal (a verde). Os asteriscos azuis representam os pontos obtidos pela PDD, correspondentes a iterações bem sucedidas.

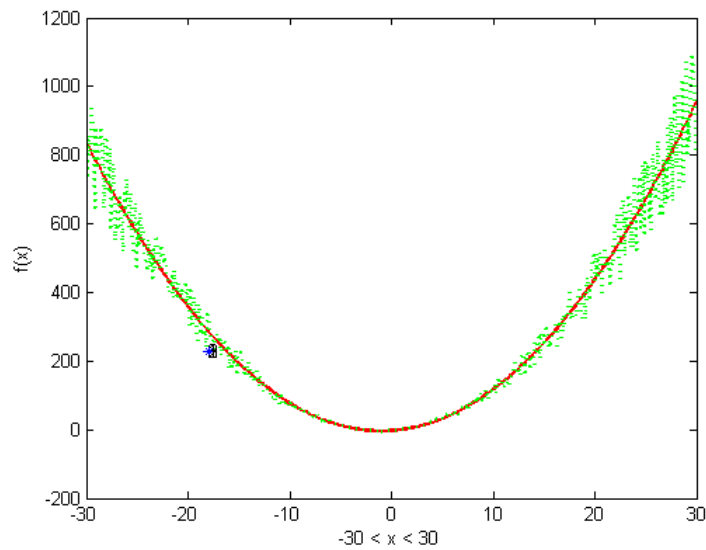


Figura 3.6: Gráfico da função f (a vermelho) perturbada com um nível de ruído de 20%, baseado em ruído Uniforme (a verde). Os asteriscos azuis representam os pontos obtidos pela PDD, correspondentes a iterações bem sucedidas.

Ao observar estes três gráficos pode-se constatar que, considerando um mesmo nível de ruído (20%), os diferentes tipos de ruído têm impactos distintos na avaliação da função, resultando em diferentes tipos de oscilações: mais regulares quando são usados polinômios de Chebyshev e menos regulares nas restantes situações. Esta falta de regularidade faz com que frequentemente se observe a incapacidade da PDD em determinar o mínimo

do problema.

Realizou-se então um estudo numérico que procura avaliar a capacidade da PDD em ser bem sucedida, considerando os três tipos de ruído numérico (baseado em polinómios de Chebyshev, ruído Uniforme e ruído Normal) e as três percentagens de ruído (5%, 10% e 20%). As Tabelas 3.1–3.3 respeitam à variação da percentagem de ruído presente na função objectivo. As Tabelas 3.4–3.6 consideram uma percentagem de ruído fixa e fazem variar o tipo de ruído a que a função está sujeita.

Considerando inicializações inteiras, em qualquer uma das seis tabelas as colunas 2 e 6 registam as inicializações mais distantes do ponto óptimo que permitem a convergência da PDD para o óptimo (o correspondente valor da função objectivo encontra-se nas colunas 3 e 7). As colunas 4 e 8 representam as inicializações mais perto do ponto óptimo para as quais a PDD falha a sua determinação, sendo o correspondente valor final da função objectivo registado nas colunas 5 e 9.

Tabela 3.1: Resultados da aplicação da PDD a problemas contaminados com ruído Uniforme, considerando três níveis de ruído.

Ruído Uniforme	x_ini	f_final	x_ini	f_final	x_ini	f_final	x_ini	f_final
5%	61	-4,200	62	3.708E+03	-70	-4,200	-71	4.595E+03
10%	32	-4,200	33	1.053E+03	-54	-4,300	-55	2.693E+03
20%	20	-4,776	21	394.268	-22	-4,761	-23	427.030

Tabela 3.2: Resultados da aplicação da PDD a problemas contaminados com ruído Normal, considerando três níveis de ruído.

Ruído Normal	x_ini	f_final	x_ini	f_final	x_ini	f_final	x_ini	f_final
5%	52	-4,600	53	2.498E+03	-44	-4,500	-45	1.603E+03
10%	19	-4,794	20	361.673	-27	-4,858	-28	553.294
20%	12	-5,222	13	115.532	-17	-5,635	-18	155.774

Tabela 3.3: Resultados da aplicação da PDD a problemas contaminados com ruído baseado em polinómios de Chebyshev, considerando três níveis de ruído.

Ruído bas. pol. Chebyshev	x_ini	f_final	x_ini	f_final	x_ini	f_final	x_ini	f_final
5%	49	-4,200	50	2.540E+03	-67	-4,200	-68	4.348E+03
10%	31	-4,400	32	1.054E+03	-32	-4,400	-33	931.426
20%	25	-4,743	26	669.072	-26	-4,743	-27	576.761

Ao analisar estas três tabelas pode-se observar que, para as três classes de ruído, o nível do ruído afecta a distância do ponto óptimo a que se pode considerar a inicialização do método, por forma a garantir que a PDD é bem sucedida. Assim, para qualquer classe de ruído, um aumento do nível de ruído presente na função objectivo faz com que a PDD necessite de um ponto inicial mais perto do ponto óptimo para conseguir escapar aos

3.2. ANÁLISE NUMÉRICA DO DESEMPENHO DA PROCURA DIRECTA DIRECCIONAL EM PROBLEMAS COM RUÍDO

óptimos locais espúrios resultantes da presença desse ruído.

Tabela 3.4: Resultados da aplicação da PDD considerando um nível de ruído de 5% e os diferentes tipos de ruído.

5%	x_ini	f_final	x_ini	f_final	x_ini	f_final	x_ini	f_final
Ruído Uniforme	61	-4,200	62	3.708E+03	-70	-4,200	-71	4.595E+03
Ruído Normal	52	-4,600	53	2.498E+03	-44	-4,500	-45	1.603E+03
Ruído bas. pol. Chebyshev	49	-4,200	50	2.540E+03	-67	-4,200	-68	4.348E+03

Tabela 3.5: Resultados da aplicação da PDD considerando um nível de ruído de 10% e os diferentes tipos de ruído.

10%	x_ini	f_final	x_ini	f_final	x_ini	f_final	x_ini	f_final
Ruído Uniforme	32	-4,200	33	1.053E+03	-54	-4,300	-55	2.693E+03
Ruído Normal	19	-4,794	20	361.673	-27	-4,858	-28	553.294
Ruído bas. pol. Chebyshev	31	-4,400	32	1.054E+03	-32	-4,400	-33	931.426

Tabela 3.6: Resultados da aplicação da PDD considerando um nível de ruído de 20% e os diferentes tipos de ruído.

20%	x_ini	f_final	x_ini	f_final	x_ini	f_final	x_ini	f_final
Ruído Uniforme	20	-4,776	21	394.268	-22	-4,761	-23	427.030
Ruído Normal	12	-5,222	13	115.532	-17	-5,635	-18	155.774
Ruído bas. pol. Chebyshev	25	-4,743	26	669.072	-26	-4,743	-27	576.761

Por outro lado, a análise das Tabelas 3.4–3.6 permite concluir que, em cada nível de ruído, o tipo de ruído afecta a distância do ponto óptimo a que se pode considerar a inicialização do método, por forma a que a PDD seja bem sucedida. As classes de ruído baseado em polinómios de Chebyshev e de ruído Uniforme permitem em geral maiores distâncias entre o ponto inicial e o ponto óptimo do que a baseada em ruído Normal.

3.2 Análise numérica do desempenho da procura directa direccional em problemas com ruído

Esta segunda abordagem numérica tem objectivos semelhantes aos da abordagem anterior mas foi considerado um conjunto de 53 problemas recolhidos da literatura, cada um permitindo dois tipos de inicializações distintas (longe e perto do ponto óptimo, respectivamente). O código que define cada uma das funções objectivo pode obter-se em

<http://www.mcs.anl.gov/~more/dfo>.

Uma descrição detalhada de cada problema encontra-se em [36]. À semelhança da Secção 3.1, procurou-se ganhar sensibilidade acerca da capacidade da PDD em determinar

a solução dos problemas perturbados com diferentes tipos/níveis de ruído, o que motivará eventualmente a sua adaptação a este tipo de optimização.

3.2.1 A classe de problemas

Na Tabela 3.7 regista-se o número de problemas para cada uma das dimensões, n_p , consideradas no conjunto teste ($2 \leq n_p \leq 12$).

Tabela 3.7: Número de problemas para cada uma das dimensões, n_p , consideradas no conjunto teste ($2 \leq n_p \leq 12$).

Dimensão do problema (n_p)	2	3	4	5	6	7	8	9	10	11	12
Número de problemas	5	6	5	4	4	5	6	5	4	4	5

Cada um destes problemas pode ser combinado com os três tipos de ruído considerados na Secção 3.1: baseado em polinómios de Chebyshev [36], Uniforme [36] ou Normal [33]. Para cada classe de problemas, à semelhança do que foi feito na Secção 3.1, analisaram-se três níveis de ruído distintos: 5%, 10% e 20%.

Partindo de uma classe de problemas suaves, definida por:

$$f(x) = \sum_{k=1}^m f_k^2(x), \quad (3.6)$$

a classe de problemas com ruído Uniforme [36] é dada por

$$\tilde{f}(x) = (1 + U) \sum_{k=1}^m f_k^2(x), \quad (3.7)$$

em que U é uma variável com distribuição uniforme no intervalo $[-\varepsilon, \varepsilon]$, sendo ε o nível de ruído considerado.

A classe de problemas com ruído baseado em polinómios de Chebyshev [36] é definida por:

$$\tilde{f}(x) = (1 + \varepsilon \alpha(x)) \sum_{k=1}^m f_k^2(x), \quad (3.8)$$

em que novamente ε representa a percentagem de ruído considerada, sendo $\alpha(x)$ e α_0 definidos de acordo com as equações 3.3 e 3.4.

Finalmente, a classe de problemas com ruído Normal [33] é definida por:

$$\tilde{f}(x) = (1 + N) \sum_{k=1}^m f_k^2(x), \quad (3.9)$$

3.2. ANÁLISE NUMÉRICA DO DESEMPENHO DA PROCURA DIRECTA DIRECCIONAL EM PROBLEMAS COM RUÍDO

com N uma variável com distribuição Normal de média $\mu = 0$ e variância $\sigma^2 = \varepsilon^2$, sendo ε a percentagem de ruído considerada.

Dos 53 problemas descritos, apenas em 16 é sugerido utilizar inicializações longe do ponto óptimo, considerando um factor multiplicativo que afecte a inicialização proposta na literatura [36]. Na Tabela 3.8 representam-se as dimensões destes 16 problemas (segunda coluna) e os valores óptimos reportados na literatura [37], sem contaminação de ruído (terceira coluna). Pode-se observar que o problema 8 apresenta dois óptimos locais enquanto que para os problemas 14, 15 e 16 não foi encontrado registo dos respectivos valores óptimos. A quarta coluna apresenta o melhor ponto obtido pela procura directa direccionada considerando os valores iniciais propostos na literatura. A quinta coluna apresenta os melhores valores obtidos pela PDD considerando os valores iniciais propostos pela literatura multiplicados por 50, por forma a considerar inicializações longe do ponto óptimo.

Os resultados numéricos obtidos permitem concluir que, considerando um orçamento computacional de 1500 cálculos (razoável para problemas desta classe, com este tipo de dimensões), a PDD é, em geral, bem sucedida quando a inicialização considerada é a sugerida na literatura. Considerando o factor de 50 na inicialização, a PDD apresenta alguns casos de terminação prematura, resultante do critério de paragem considerado. Ainda assim, tendo em conta os resultados obtidos na Secção 3.1, onde a inicialização perto do óptimo permitia o sucesso da PDD, optou-se por considerar nos restantes testes computacionais o factor de 50 na inicialização.

Tabela 3.8: Dimensões, valores óptimos encontrados na literatura e valores finais obtidos pela PDD considerando inicializações perto ou longe do óptimo.

Problemas	Dimensão do problema (n_p)	Óptimo da Literatura	Valor final PDD inic. perto	Valor final PDD inic. longe
1	9	3,600E+01	3,600E+01	3,600E+01
2	7	8,380E+00	8,380E+00	8,380E+00
3	7	9,806E+00	9,881E+00	9,881E+00
4	2	0,000E+00	8,183E-02	3,797E+00
5	3	0,000E+00	0,000E+00	2,822E-03
6	4	0,000E+00	6,195E-06	8,082E+03
7	2	4,898E+01	4,898E+01	4,898E+01
8	3	8,215E-03 1,743E+01	1,508E-02	1,743E+01
9	6	2,288E-03	2,280E-01	3,811E+01
10	9	1,400E-06	1,722E+00	6,233E+01
11	12	4,722E-10	2,388E+00	4,396E+03
12	4	8,582E+04	8,582E+04	8,597E+04
13	11	4,014E-02	1,570E-01	2,631E+01
14	5	-	5,630E-07	2,496E-04
15	12	-	1,958E+05	3,886E+10
16	8	-	1,249E-01	1,629E+11

3.2.2 Perfis de dados

A comparação do desempenho numérico de algoritmos pode ser difícil quando estes são em grande número ou quando o conjunto teste apresenta um número elevado de problemas. Em optimização sem recurso a derivadas é frequente a utilização de perfis de dados [36] nesta comparação.

Estes perfis procuram medir a percentagem de problemas resolvidos por cada algoritmo, dentro de um certo orçamento computacional (medido em número de cálculos da função objectivo). Um problema é considerado resolvido quando satisfaz a seguinte condição:

$$f(x_0) - f(x) \geq (1 - \tau)(f(x_0) - f_L),$$

em que $\tau > 0$ representa o nível de tolerância considerado para a elaboração do perfil de dados, x_0 a inicialização considerada para o problema e f_L o menor valor obtido para a função objectivo por qualquer um dos algoritmos considerados (idealmente o óptimo do problema).

Este teste de convergência significa que um problema é considerado como solucionado por um dado algoritmo se a redução $f(x_0) - f(x)$ obtida por esse algoritmo for igual a pelo menos $(1 - \tau)$ vezes a melhor redução $f(x_0) - f_L$ obtida por qualquer um dos algoritmos considerados na resolução do problema em causa.

Seja S o conjunto de todos os algoritmos considerados e P o conjunto de todos os problemas testados. A definição de um perfil de dados para o algoritmo $s \in S$ consiste na representação gráfica da curva $d_s(\alpha) = \frac{1}{|P|} \left| \left\{ p \in P : \frac{t_{p,s}}{n_p + 1} \leq \alpha \right\} \right|$, onde n_p é a dimensão do problema $p \in P$ e $t_{p,s} > 0$ uma medida de desempenho para cada $p \in P$ e $s \in S$ (por exemplo, o número de cálculos necessários para o algoritmo s resolver o problema p). Esta medida é dividida por $n_p + 1$ normalizando $t_{p,s}$, por forma a que a dimensão do problema não influencie o desempenho do algoritmo.

Assim, pela definição anterior, $d_s(\alpha)$ pode ser interpretado como a percentagem de problemas resolvidos pelo algoritmo s , para um máximo de $\alpha(n_p + 1)$ cálculos da função objectivo [36].

3.2.3 Experiência numérica

Os perfis de dados descritos na secção anterior são utilizados para apresentar os resultados da PDD nas classes de problemas descritas na Secção 3.2.1, considerando os 16 problemas para os quais é sugerido utilizar inicializações longe do ponto óptimo. Em cada perfil de dados a PDD é testada no conjunto de problemas sem ruído (*SIM_MD*) e no conjunto de problemas contaminado com ruído (*SIM_MD_PR*). Os perfis foram elaborados considerando uma tolerância $\tau = 10^{-3}$.

3.2. ANÁLISE NUMÉRICA DO DESEMPENHO DA PROCURA DIRECTA DIRECCIONAL EM PROBLEMAS COM RUÍDO

Dado que os ruídos Normal e Uniforme são gerados aleatoriamente, por forma a diminuir a dependência das sementes consideradas na sua geração, fixou-se um conjunto de várias sementes, também elas geradas de forma aleatória. Para tal, na função do Matlab *rng* optou-se pelo gerador ‘*twister*’. No total utilizaram-se 10 sementes, fazendo com que, para estes dois tipos de ruído, o número de problemas a testar seja 160, ao contrário do ruído baseado em polinómios de Chebyshev que mantém os 16 problemas descritos, uma vez que se trata de ruído determinístico. Nas Figuras 3.7–3.9 encontram-se os resultados obtidos com a PDD para cada uma das classes de problemas.

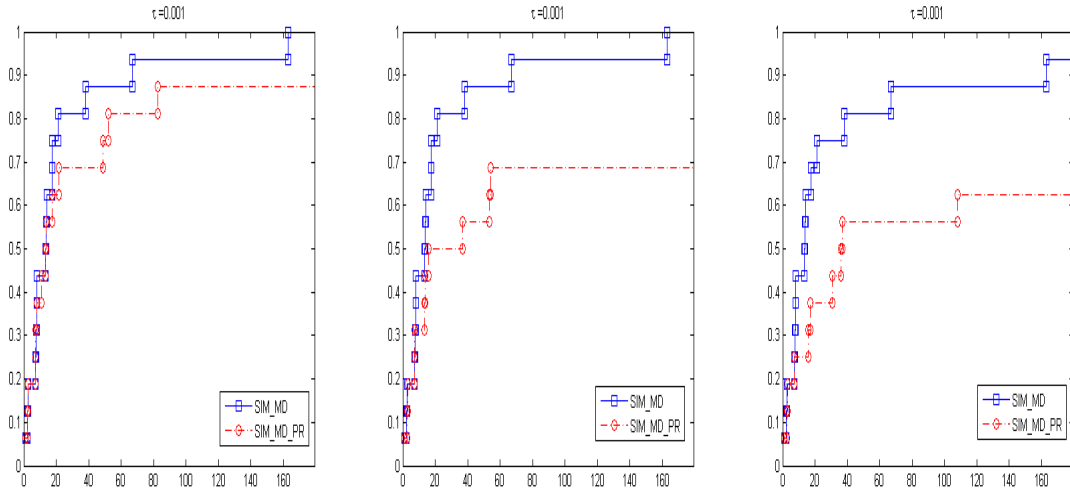


Figura 3.7: Perfis de dados para o conjunto de problemas perturbados com ruído baseado em polinómios de Chebyshev, para os níveis de ruído de 5%, 10% e 20%, respectivamente.

Ao analisar os três perfis de dados referentes ao ruído baseado em polinómios de Chebyshev (ver a Figura 3.7) pode-se constatar que o nível de ruído presente na função objectivo afecta o desempenho da PDD. À medida que se aumenta este nível de ruído, o desempenho da PDD piora, sendo que para 20% de ruído apenas cerca de 61% dos problemas são resolvidos. Nota-se ainda que, quando são permitidos até $20(n_p + 1)$ cálculos para a função objectivo, a PDD parece competitiva, mesmo na presença de ruído.

Quando o ruído é baseado na distribuição Uniforme (ver a Figura 3.8), tal como no caso da Secção 3.1, são extraídas conclusões idênticas às obtidas para o tipo de ruído anterior. Considerando um nível de ruído de 20% constata-se que a PDD apenas resolve cerca de 64% dos problemas, sendo competitiva para orçamentos de $20(n_p + 1)$ cálculos, independentemente do nível de ruído considerado.

No que respeita à perturbação resultante de ruído baseado na distribuição Normal (ver a Figura 3.9), a situação agrava-se. Com a presença de 5% de ruído, a PDD apenas consegue resolver cerca de 73% dos problemas, reduzindo-se esta percentagem para 43%

CAPÍTULO 3. ANÁLISE DE DESEMPENHO DA PROCURA DIRECTA DIRECCIONAL EM OPTIMIZAÇÃO DE PROBLEMAS COM RUÍDO

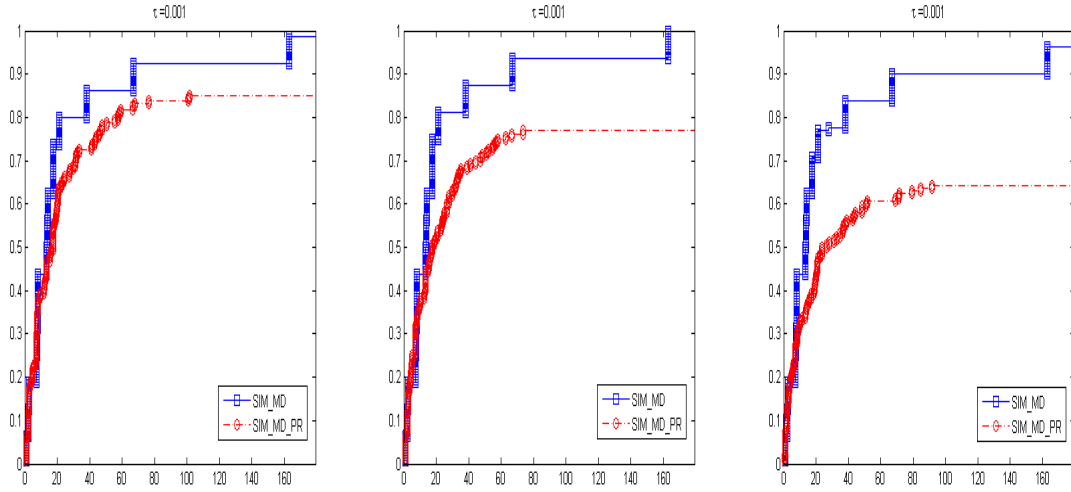


Figura 3.8: Perfis de dados para o conjunto de problemas perturbados com ruído Uniforme, para os níveis de ruído de 5%, 10% e 20%, respectivamente.

no caso da presença de 20% de ruído gerado segundo a distribuição Normal. O limite ao número máximo de cálculos permitido para a função objectivo, dentro do qual a PDD é competitiva na presença de ruído é claramente mais baixo, por comparação com os dois tipos de ruído anteriormente considerados.

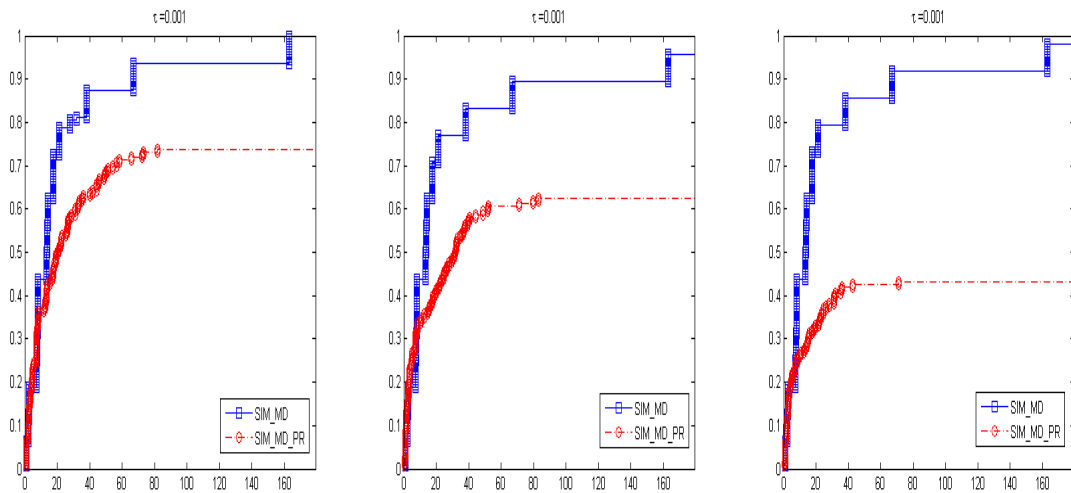


Figura 3.9: Perfis de dados para o conjunto de problemas perturbados com ruído Normal para os níveis de ruído de 5%, 10% e 20%, respectivamente.

Os estudos numéricos conduzidos nesta secção justificam a necessidade da definição de estratégias que permitam aumentar o desempenho da PDD quando aplicada a problemas perturbados com ruído, desde que não haja a certeza de se estar a considerar uma inicialização perto do óptimo.

ESTRATÉGIAS NÃO MONÓTONAS

As estratégias não monótonas foram primeiramente propostas em métodos de procura unidireccional na presença de derivadas [24]. No trabalho mencionado, com o objectivo de acelerar a convergência em estádios intermédios do processo de optimização, facilitando a aceitação de comprimentos de passos unitários, os autores propuseram uma modificação do método de Newton em que a aceitação de um ponto como uma nova iterada implica que o correspondente valor da função objectivo satisfaça a condição de Armijo [3], não em relação ao valor da função objectivo da iterada corrente mas em relação ao valor máximo da função objectivo, considerando um número previamente fixo de iteradas anteriores.

Esta técnica de procura unidireccional pode ser vista como uma generalização da regra de Armijo, permitindo a convergência do método, mas relaxando o comportamento monótono. Por outras palavras, neste tipo de abordagem procura-se melhorar o valor da função objectivo considerando um histórico de iterações, o que pode conduzir a acréscimos no valor da função objectivo entre iterações consecutivas (daí a terminologia não monótona).

Para estes autores, um novo ponto $x_{k+1} = x_k + \alpha_k d_k$ é bem sucedido se

$$f(x_{k+1}) \leq \max_{0 \leq j \leq m(k)} \{f(x_{k-j})\} + \gamma \alpha_k g_k^\top d_k,$$

onde $\alpha_k = \sigma^h a$, com $a > 0$, $\sigma \in (0, 1)$, $\gamma \in (0, 1)$, $d_k \neq 0$ representa uma direcção, g_k o gradiente de f em x_k e h o primeiro inteiro não negativo que satisfaz a condição anterior, com $m(0) = 0$, $0 \leq m(k) \leq \min\{m(k-1) + 1, M\}$ para $k \geq 1$. O histórico de iterações considerado é definido por $\{f(x_{l(k)})\}$, em que $k - m(k) \leq l(k) \leq k$, com M um inteiro não negativo.

Estas estratégias foram mais tarde introduzidas por Diniz-Ehrhardt, Martínez e Raydan em métodos de procura unidireccional sem recurso a derivadas [19]. O método que propõem combina e estende a técnica de procura unidireccional não monótona anteriormente relatada e a de Li e Fukushima [35]. No contexto da solução de sistemas de

equações não lineares $F(x) = 0$, Li e Fukushima propuseram um método de quasi-Newton baseado em procura unidireccional, modificando a condição de aceitação de novos pontos por forma a que fosse independente do cálculo de derivadas.

Seja $\{\eta_k\}_{k \in \mathbb{N}_0}$ uma sucessão de números positivos tal que $\sum_{k=0}^{\infty} \eta_k < \infty$. Considere-se a constante $\sigma_1 > 0$, d_k uma direcção e $\lambda > 0$ um comprimento de passo. Em cada iteração será satisfeita a condição

$$\|F(x_k + \lambda d_k)\| \leq \|F(x_k)\| - \sigma_1 \|\lambda d_k\|^2 + \eta_k \|F(x_k)\|$$

o que garante que

$$\|F(x_k + \lambda d_k)\| \leq (1 + \eta_k) \|F(x_k)\|.$$

Justifica-se assim a designação atribuída a esta procura unidireccional– *procura unidireccional de norma aproximadamente descendente*. Apesar deste método tolerar direcções não descendentes, o seu comportamento não monótono é insuficiente.

Em [19], os autores procuram ultrapassar esta limitação combinando a abordagem de Li e Fukushima com a apresentada em [24], que garante um comportamento suficientemente não monótono, mas exige o conhecimento de derivadas e o uso de direcções de descida para garantir a convergência.

Para tal, os autores consideram as sucessões de números positivos $\{\eta_k\}_{k \in \mathbb{N}_0}$ e $\{\beta_k\}_{k \in \mathbb{N}_0}$ com $\sum_{k=0}^{\infty} \eta_k < \infty$ e $\{\beta_k\}_{k \in \mathbb{N}_0}$ uma sucessão limitada tal que para todo o subconjunto infinito de índices $K \subseteq \mathbb{N}$,

$$\lim_{k \in K} \beta_k = 0 \Rightarrow \lim_{k \in K} \nabla f(x_k) = 0.$$

Um ponto será aceite como uma nova iterada se satisfaz a condição

$$f(x_k + \alpha d_k) \leq \max_{\max\{k-M+1, 0\} \leq j \leq k} \{f(x_j)\} - \alpha_k^2 \beta_k + \eta_k.$$

Esta técnica de procura unidireccional é testada numericamente, considerando quer conjuntos de direcções geradas aleatoriamente em cada iteração, quer direcções baseadas em aproximações ao gradiente ou direcções de quasi-Newton que utilizam a actualização SR1 [25]. Em qualquer dos casos, as propriedades de convergência dos algoritmos resultam em exclusivo da técnica de procura unidireccional considerada.

Algumas variantes de estratégias não monótonas exigem o decréscimo não do máximo do histórico de iterações mas da sua média [46]. Esta abordagem foi considerada em métodos de procura unidireccional sem recurso a derivadas [39], incluindo a resolução de sistemas de equações não lineares de grande escala [11].

Em [23] é proposta uma classe geral de métodos não monótonos de procura unidireccional, que consideram em cada iteração diferentes conjuntos de direcções (direcções coordenadas, gradientes simpléticos, rotações das direcções coordenadas). Novamente, a convergência resulta da técnica de procura unidireccional considerada.

ESTRATÉGIAS NÃO MONÓTONAS EM PROCURA DIRECTA DIRECCIONAL NA OPTIMIZAÇÃO DE PROBLEMAS COM RUÍDO

Como foi mencionado no capítulo anterior na abordagem não monótona não se procura melhorar o valor da função objectivo entre iterações consecutivas mas sim ao longo de um histórico de iterações. Esta abordagem, num contexto em que a função objectivo está contaminada com ruído numérico, pode permitir escapar a mínimos locais espúrios, resultantes do ruído. Neste capítulo serão propostos dois algoritmos de PDD baseados em estratégias não monótonas e será analisada a sua convergência.

5.1 Descrição das variantes algorítmicas

Qualquer um dos algoritmos propostos é um algoritmo de procura directa direcciona, que segue a estrutura descrita na Figura 2.1, mas equipado com uma estratégia não monótona para a aceitação de novos pontos, tendo sido consideradas duas variantes não monótonas distintas.

Seguindo [19, 33], a primeira variante não monótona designada por *NMD* começa por fixar $M \in \mathbb{N}$ que representa o comprimento do histórico. Em cada iteração k da procura directa direcciona, quer no passo de procura, quer no passo de sondagem, um novo ponto, x_j , será aceite se $f(x_j) < \max\{f(x_k), f(x_{k-1}), \dots, f(x_{k-M+1})\}$. Nesta variante, o histórico de valores da função objectivo que serão utilizados para comparação com um novo ponto será actualizado removendo-se o ponto mais antigo.

Nesta estratégia, um ponto correspondente ao melhor valor da função objectivo, por ser o mais antigo, pode vir a ser removido do histórico de pontos usados para comparação.

Esta foi a motivação para a segunda variante não monótona a considerar, designada por *BWNMD*.

Tal como na variante anterior, em cada iteração k da procura directa direccional um novo ponto, x_j , será aceite se $f(x_j) < \max\{f(x_k), f(x_{k-1}), \dots, f(x_{k-M+1})\}$. Contudo, o histórico guarda agora os últimos M melhores valores da função objectivo (ao invés de apenas os últimos M valores). Quer isto dizer que, em caso de sucesso, o ponto a ser removido do histórico deverá corresponder ao valor mais elevado da função objectivo.

Ao contrário da anterior, esta variante é mais exigente na aceitação de novos pontos (enfraquecendo assim o seu cariz não monótono), podendo ser pensada como uma versão intermédia entre a variante não monótona anterior (*NMD*) e uma estratégia monótona.

Uma descrição formal da construção do histórico de pontos e da sua actualização encontra-se na Figura 5.1.

Seja $M \in \mathbb{N}$ e F_k o histórico de valores da função objectivo a utilizar na iteração k , com $|F_k| \leq M$ e $F_0 = \{f(x_0)\}$.
 Se a iteração k é uma iteração sem sucesso $F_{k+1} = F_k$.
 Suponha-se que a iteração k é uma iteração bem sucedida:
 Se $|F_k| < M$ então $F_{k+1} = F_k \cup \{f(x_{k+1})\}$.
 Se $|F_k| = M$ então:
 No caso de se ter considerado a variante *NMD* virá $F_{k+1} = F_k \setminus \{f(x_{k-M+1})\} \cup \{f(x_{k+1})\}$.
 Caso tenha sido considerada a estratégia *BWNMD* ter-se-á $F_{k+1} = F_k \setminus \{f(x_j)\} \cup \{f(x_{k+1})\}$ com $f(x_j) = \max F_k$.

Figura 5.1: Construção e actualização do histórico de pontos nas estratégias não monótonas.

Qualquer uma das variantes não monótonas foi considerada em conjunto com as duas estratégias de globalização de PDD descritas na Secção 2.4. Na Figura 5.2 encontra-se uma descrição detalhada da estrutura algorítmica considerada. Para tal, define-se a função $\bar{\rho}$ que coincidirá com a função ρ , no caso da estratégia de globalização se basear na imposição de uma condição de decréscimo suficiente, ou será a função nula, quando a globalização é baseada no uso de grelhas racionais.

$$\bar{\rho}(\cdot) = \begin{cases} 0, & \text{se a estratégia de globalização usa grelhas racionais} \\ \rho(\cdot), & \text{se a estratégia de globalização usa decréscimo suficiente} \end{cases} \quad (5.1)$$

Inicialização

Considerar $x_0 \in \mathbb{R}^n$ com $f(x_0) < +\infty$, um conjunto de conjuntos geradores positivos ou de bases positivas D de \mathbb{R}^n e um comprimento inicial de passo $\Delta_0 > 0$. Considerar ainda $\beta \geq 1$ e $0 < \mu < 1$, utilizados na actualização do comprimento do passo.

Inicializar o histórico de pontos $F_0 = \{f(x_0)\}$. Seja $k = 0$.

Passo de Procura

Avaliar a função objectivo num número finito de pontos, tentando encontrar um ponto $x \in \mathbb{R}^n$ que satisfaça $f(x) < \max F_k - \bar{\rho}(\Delta_k)$. Se tal for possível, definir a iteração como bem sucedida, considerar $x_{k+1} = x$ e omitir o passo de sondagem. Caso contrário, realizar o passo de sondagem.

Passo de Sondagem

Escolher $D_k \in D$ e considerar o conjunto de sondagem $P_k = \{x_k + \Delta_k d : d \in D_k\}$. Avaliar f em P_k .

Se existir $x_k + \Delta_k d \in P_k$ tal que $f(x_k + \Delta_k d) < \max F_k - \bar{\rho}(\Delta_k)$, considerar $x_{k+1} = x_k + \Delta_k d$ e definir a iteração como bem sucedida. Caso contrário, considerar $x_{k+1} = x_k$ e definir a iteração como um insucesso.

Actualização do Comprimento do Passo e do Histórico de Pontos

Se a iteração foi bem sucedida, fazer $\Delta_{k+1} = \beta \Delta_k$. Caso contrário, considerar $\Delta_{k+1} = \mu \Delta_k$. Actualizar o histórico de pontos de F_k de acordo com a descrição algorítmica da Figura 5.1.

Incrementar o contador de iterações, fazendo $k = k + 1$ e voltar ao passo de procura.

Figura 5.2: Um algoritmo de procura directa direcciona, considerando estratégias não monótonas.

No caso de ser considerada a estratégia de globalização baseada em grelhas racionais designar-se-ão as variantes algorítmicas como *SIM_NMD* e *SIM_BWNMD*. Caso a estratégia de globalização seja baseada na imposição de uma condição de decréscimo suficiente, as variantes algorítmicas serão designadas como *SUF_NMD* e *SUF_BWNMD*.

5.2 Análise de convergência

A análise de convergência das variantes de PDD não monótonas segue a abordagem descrita na Secção 2.6 para o caso monótono. Assim, começa-se por provar a existência de uma subsucessão de comprimentos de passo que converge para zero, garantindo-se a convergência de subsucessões de pontos geradas pelo algoritmo.

Teorema 5.2.1. *Sob a Hipótese 2.6.1, um método de procura directa direccional usando qualquer uma das variantes não monótonas SIM_NMD e SIM_BWNMD, que consideram estratégias de globalização baseadas em grelhas racionais, gera uma sucessão de comprimentos de passo tal que,*

$$\liminf_{k \rightarrow +\infty} \Delta_k = 0.$$

Demonstração. Suponha-se que existe um $\Delta^* > 0$, tal que $\Delta_k > \Delta^*$, $\forall k \in \mathbb{N}$. Todos os pontos gerados pelo algoritmo pertencem a uma grelha racional implícita que constitui um reticulado [4]. A intersecção de um reticulado com um conjunto limitado é finita, pelo que há um número finito de pontos que podem corresponder a novas iteradas. Ao fim de M iterações bem sucedidas necessariamente o algoritmo aceita um novo ponto. Logo, há um número finito de pontos que se podem aceitar como novas iteradas em iterações bem-sucedidas, o que significa que há finitos sucessos e infinitos insucessos. Em cada iteração mal sucedida o comprimento de passo é reduzido em $\mu > 0$, o que conduz a um absurdo. □

O mesmo tipo de resultado pode ser obtido nas variantes SUF_NMD e SUF_BWNMD , quando a estratégia de globalização é baseada na imposição de uma condição de decréscimo suficiente.

Teorema 5.2.2. *Sob a Hipótese 2.6.2, um método de procura directa direccional, usando qualquer uma das variantes não monótonas SUF_NMD e SUF_BWNMD , que consideram estratégias de globalização baseadas na imposição de uma condição de decréscimo suficiente, gera uma sucessão de comprimentos de passo tal que,*

$$\liminf_{k \rightarrow +\infty} \Delta_k = 0.$$

Demonstração. Suponha-se que existe um $\Delta^* > 0$, tal que $\Delta_k > \Delta^*$, $\forall k \in \mathbb{N}$. Recorrendo às propriedades da função $\rho(\cdot)$, tem-se $\rho(\Delta_k) \geq \rho(\Delta^*) > 0$. Após $M' \geq M$ sucessos pode-se afirmar que $\max F_{k+M'} < \max F_k - \rho(\Delta_{k+M'})$. Assim, $\max F_{k+M'} - \max F_k < -\rho(\Delta_{k+M'}) \leq -\rho(\Delta^*) < 0$, o que leva a concluir que $\max F_k \rightarrow -\infty$. Desta forma, f não seria minorada, contradizendo a Hipótese 2.6.2. □

À semelhança do que sucede no caso monótono, considerando a Hipótese 2.6.1, visto que

$$\liminf_{k \rightarrow +\infty} \Delta_k = 0$$

e o comprimento de passo é reduzido apenas em insucessos, pode-se garantir a existência de uma subsucessão refinada $\{x_k\}_{k \in K}$ convergente. Ou seja, pode-se garantir a existência de uma subsucessão $\{x_k\}_{k \in K}$, correspondente a iterações sem sucesso, tal que

$$\lim_{k \in K} \Delta_k = 0.$$

Os resultados de convergência serão estabelecidos analisando o comportamento do algoritmo no ponto limite x^* , da subsucessão refinada.

Teorema 5.2.3. *Seja x^* o ponto limite de uma subsucessão refinada convergente gerada com um método de PDD, considerando uma estratégia não monótona e f continua à Lipschitz numa vizinhança de x^* . Suponha-se ainda que f é avaliada num número infinito de vezes na direcção d usada na definição dos conjuntos de sondagem da subsucessão refinada. Então a derivada direccionada generalizada de Clarke da função f definida em x^* , na direcção d , é não negativa, ou seja,*

$$f^o(x^*; d) \geq 0.$$

Demonstração. Utilizando a definição de derivada generalizada de Clarke vem

$$f^o(x^*; d) = \limsup_{\substack{x \rightarrow x^* \\ t \downarrow 0}} \frac{f(x + td) - f(x)}{t} \geq \lim_{k \in K} \frac{f(x_k + \Delta_k d) - f(x_k)}{\Delta_k} \quad (5.2)$$

onde a última desigualdade resulta de $\{x_k\}_{k \in K}$ ser uma subsucessão refinada convergente para x^* .

Numa estratégia de globalização baseada em grelhas racionais, sendo $\{x_k\}_{k \in K}$ uma subsucessão refinada, logo correspondente a iterações mal sucedidas, tem-se

$$f(x_k + \Delta_k d) \geq \max F_k \geq f(x_k).$$

Esta desigualdade garante a não negatividade do último quociente em (5.2) e como tal a não negatividade de $f^o(x^*; d)$.

Considere-se agora o caso em que a estratégia de globalização é baseada na imposição de uma condição de decréscimo suficiente. Sendo $\{x_k\}_{k \in K}$ uma subsucessão refinada vem

$$f(x_k + \Delta_k d) \geq \max F_k - \rho(\Delta_k) \geq f(x_k) - \rho(\Delta_k).$$

Assim, atendendo às propriedades da função $\rho(\cdot)$, vem:

$$\begin{aligned} f^o(x^*; d) &= \limsup_{\substack{x \rightarrow x^* \\ t \downarrow 0}} \frac{f(x + td) - f(x)}{t} \geq \\ \lim_{k \in K} \frac{f(x_k + \Delta_k d) - f(x_k)}{\Delta_k} &= \lim_{k \in K} \frac{f(x_k + \Delta_k d) - f(x_k) + \rho(\Delta_k)}{\Delta_k} - \frac{\rho(\Delta_k)}{\Delta_k} \geq 0. \end{aligned}$$

□

Tal como sucede no Teorema 2.6.4 uma versão mais robusta pode ser obtida se se assumir mais suavidade acerca da função objectivo.

Teorema 5.2.4. *Seja x^* o ponto limite de uma subsucessão refinada convergente gerada com um método de PDD, considerando uma estratégia não monótona. Se f for estritamente diferenciável em x^* , então*

$$\nabla f(x^*) = 0.$$

A demonstração ao teorema é análoga ao caso monótono, descrito na Secção 2.6.

ANÁLISE DE DESEMPENHO DAS ESTRATÉGIAS NÃO MONÓTONAS EM PDD NA OPTIMIZAÇÃO DE PROBLEMAS COM RUÍDO

Neste capítulo procura-se avaliar o desempenho numérico das variantes algorítmicas de PDD não monótona propostas no capítulo anterior. Em particular, pretende-se estabelecer a eventual vantagem da sua utilização por comparação com algoritmos monótonos de PDD.

Para tal considerou-se o conjunto de 16 problemas teste descrito na Secção 3.2.1, que permitem inicializações longe do ponto óptimo. Para cada problema considerou-se os três tipos de ruído anteriormente descritos (ruído baseado em polinómios de Chebyshev, ruído Normal e ruído Uniforme) e três níveis de ruído (5%, 10% e 20%).

As implementações algorítmicas foram feitas em Matlab e seguem as descrições das Figuras 2.1 e 5.2, para as estratégias monótona e não monótona, respectivamente. Em qualquer dos casos, utilizou-se o conjunto gerador positivo $[e_n, -e_n, I_n, -I_n]$, em que e_n é um vector de uns com dimensão n e I_n representa a matriz identidade. Considerou-se ainda um comprimento de passo inicial $\Delta_0 = 1$ e os parâmetros $\beta = 2$ e $\mu = \frac{1}{2}$ na actualização do comprimento de passo. O passo de sondagem é efectuado de forma oportunista, tal como descrito na Secção 2.3. No caso da estratégia de globalização ser baseada na imposição de uma condição de decréscimo suficiente, utilizou-se a função $\rho(t) = t^2$.

Relativamente ao comprimento máximo permitido para o histórico de pontos das variantes não monótonas, foram testados diversos valores, reportando-se apenas as variantes correspondentes ao menor comprimento do histórico, que permitiu obter os melhores resultados. Novamente, como critério de paragem, considerou-se um máximo de 1500 cálculos para a função objectivo.

Os resultados serão reportados recorrendo aos perfis de dados descritos na Secção

3.2.2. Resultados numéricos mais detalhados encontram-se nas tabelas do Apêndice A.

6.1 Estratégias não monótonas

6.1.1 Ruído baseado em polinômios de Chebyshev

Analisando os três perfis de dados da Figura 6.1 constata-se que há uma clara vantagem em utilizar estratégias não monótonas, independentemente do nível de ruído considerado, desde que o orçamento computacional seja moderado.

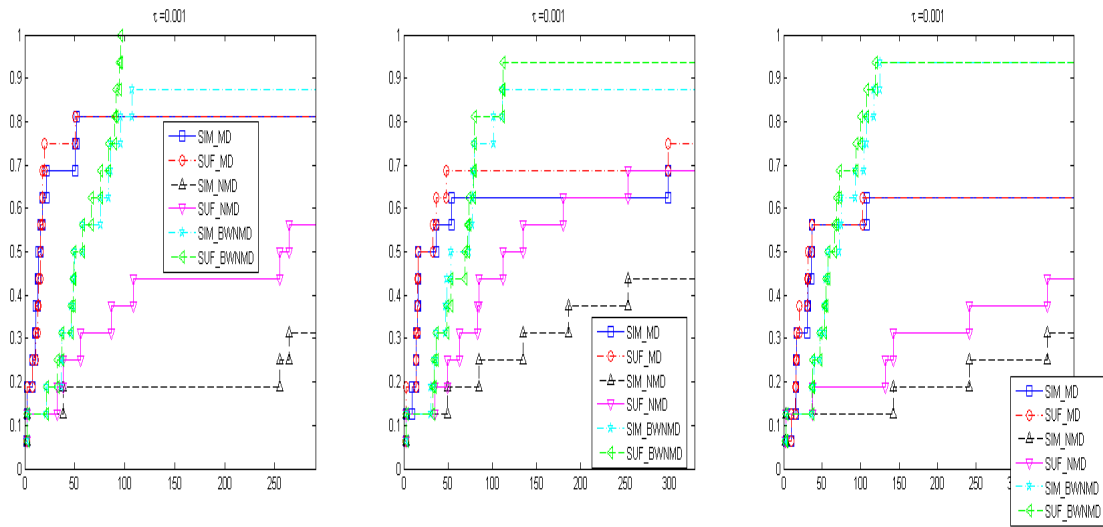


Figura 6.1: Perfis de dados para o conjunto de problemas perturbados com ruído baseado em polinômios de Chebyshev, para os níveis de ruído de 5%, 10% e 20%, com o comprimento do histórico de 5, 4 e 6 respectivamente.

Para um nível de 5% de ruído observa-se que a estratégia não monótona *SUF_BWNMD* resolve 100% dos problemas, enquanto que as duas estratégias monótonas resolvem apenas 81%. Observa-se ainda que a estratégia *SIM_BWNMD* apresenta igualmente um desempenho superior, quando comparada com as estratégias monótonas, quando o orçamento computacional (em número de cálculos da função objectivo) é moderado. Para orçamentos computacionais muito reduzidos, as estratégias monótonas são mais vantajosas.

Aumentando o nível de ruído para 10% constata-se que a estratégia não monótona que resolve mais problemas continua a ser a *SUF_BWNMD*, sendo que resolve cerca de 94% de problemas. Observa-se ainda que a estratégia *SIM_BWNMD* apresenta igualmente um bom desempenho, sendo superior ao de qualquer estratégia monótona. Para este nível de ruído verifica-se que há um decréscimo da percentagem de problemas resolvidos pelas duas estratégias monótonas, sendo este decréscimo mais evidente na estratégia *SIM_MD*.

Analisando o perfil de dados para um nível de 20% de ruído verifica-se que a estratégia que resolve mais problemas é a *BWNMD* independentemente da estratégia de globalização considerada. Observa-se ainda que em relação ao nível de ruído anterior há

um ligeiro decréscimo no desempenho da melhor estratégia monótona, que agora resolve cerca de 63% dos problemas.

Pode-se concluir assim que para este tipo de ruído há uma grande vantagem em utilizar estratégias não monótonas, sendo que a melhor estratégia é a *BWNMD* apresentando uma ligeira vantagem em utilizar uma estratégia de globalização baseada na imposição de uma condição de decréscimo suficiente. Apenas orçamentos computacionais reduzidos justificam o uso de estratégias monótonas.

6.1.2 Ruído Normal

Os três perfis de dados da Figura 6.2 mostram que há uma grande vantagem em utilizar estratégias não monótonas quando o orçamento computacional (em número de cálculos da função objectivo) é moderado. Para orçamentos computacionais reduzidos, as estratégias monótonas são mais vantajosas.

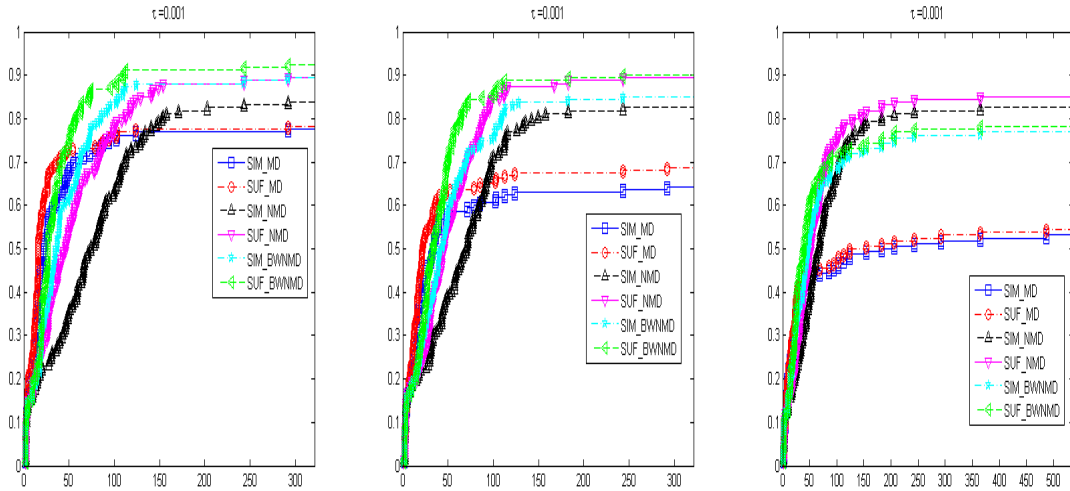


Figura 6.2: Perfis de dados para o conjunto de problemas perturbados com ruído Normal, para os níveis de ruído de 5%, 10% e 20%, respectivamente, em qualquer dos casos com o comprimento do histórico igual a 2.

Analisando o perfil de dados para um nível de 5% de ruído constata-se que a estratégia que resolve mais problemas é a estratégia não monótona *SUF_BWNMD*, resolvendo cerca de 94% de problemas, enquanto que as estratégias monótonas resolvem apenas 78%. Observa-se que as outras estratégias não monótonas apresentam igualmente um bom desempenho em relação às estratégias monótonas. Constata-se que para um número reduzido de cálculos as estratégias monótonas apresentam um desempenho superior às não monótonas.

Para um nível de ruído de 10% constata-se que as estratégias que apresentam um melhor desempenho continuam a ser as não monótonas. Quer a estratégia *SUF_BWNMD*, quer a estratégia *SUF_NMD* resolvem cerca de 90% dos problemas, embora a última estratégia necessite de mais cálculos para atingir esta percentagem. Em relação ao nível

de ruído anterior, observa-se que há um decréscimo de desempenho das estratégias monótonas, que ainda assim continuam a ser vantajosas para orçamentos computacionais reduzidos.

Analisando o perfil de dados para um nível de 20% de ruído constata-se que a estratégia *SUF_NMD* é agora a estratégia que apresenta o melhor desempenho, resolvendo 85% dos problemas. O facto da estratégia *SUF_BWNMD*, para este nível de ruído, deixar de ser a estratégia que apresenta melhor desempenho pode dever-se ao facto deste tipo de ruído apresentar uma maior irregularidade na amplitude das oscilações e desta estratégia promover menos o comportamento não monótono do que a *SUF_NMD*. Para este nível de ruído continua-se a verificar que há um decréscimo no desempenho das estratégias monótonas em relação aos níveis de ruído anteriores. Mesmo quando o número de cálculos é reduzido, as estratégias monótonas já não apresentam um desempenho tão bom como quando comparadas com os níveis de ruído mais baixos.

Para este tipo de ruído conclui-se que as estratégias não monótonas apresentam um desempenho muito superior às estratégias monótonas. Verifica-se que as estratégias monótonas no geral apresentam um mau desempenho, o que pode ser justificado pela irregularidade na forma como o ruído afecta os vários problemas (ver a Figura 3.5). As melhores estratégias para resolver problemas com este tipo de ruído são as estratégias não monótonas *SUF_BWNMD* e *SUF_NMD*, dependendo da presença de níveis mais baixos ou elevados de ruído.

6.1.3 Ruído Uniforme

Analisando os três perfis de dados da Figura 6.3 novamente se observa que há vantagem em utilizar estratégias não monótonas para este tipo de ruído, quando o número de cálculos disponível para a função objectivo é moderado. As estratégias monótonas (em particular a estratégia *SUF_MD*) continuam a apresentar vantagem para orçamentos computacionais reduzidos.

Para um nível de 5% de ruído verifica-se que a estratégia não monótona *SUF_BWNMD* resolve cerca de 96% de problemas, enquanto que as estratégias monótonas resolvem apenas 85%. Observa-se também que a única estratégia não monótona que não apresenta um bom desempenho é a *SIM_NMD*.

Aumentando o nível de ruído para 10% constata-se que a estratégia não monótona *SUF_BWNMD* continua a ser a mais promissora, resolvendo cerca de 95% de problemas. Verifica-se que, em comparação com nível de ruído anterior, a estratégia *SIM_NMD* consegue ser ligeiramente mais competitiva.

Analisando o perfil de dados para um nível de 20% de ruído verifica-se que a estratégia que resolve mais problemas é novamente a *SUF_BWNMD*, resolvendo cerca de 90% dos problemas. Observa-se que há um ligeiro decréscimo no desempenho das estratégias monótonas em relação aos níveis de ruído anteriores. Constata-se que, quando o número de cálculos é reduzido, as estratégias não monótonas apresentam um desempenho mais

próximo do das estratégias monótonas, por comparação com os níveis de ruído mais baixos.

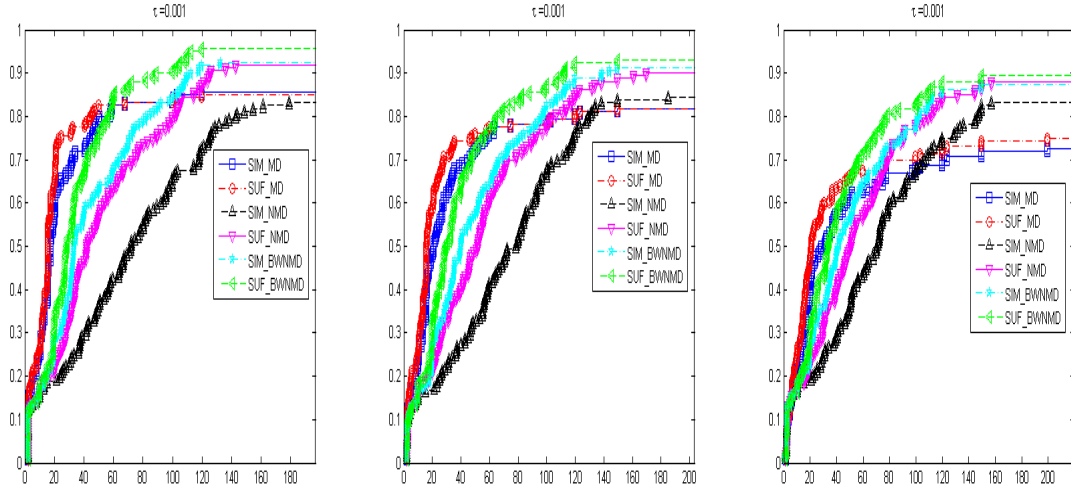


Figura 6.3: Perfis de dados para o conjunto de problemas perturbados com ruído Uniforme, para os níveis de ruído de 5%, 10% e 20%, respectivamente, em qualquer dos casos com o comprimento do histórico igual a 2.

Pode-se concluir assim que para este tipo de ruído há uma clara vantagem em utilizar estratégias não monótonas, sendo que a melhor estratégia é a *BWNMD*, principalmente se se considerar a estratégia de globalização que se baseia na imposição de decréscimo suficiente. As estratégias monótonas apresentam um desempenho superior às estratégias não monótonas apenas quando é considerado um número reduzido de cálculos, devendo nessa caso optar-se pela estratégia *SUF_MD*, que também usa como estratégia de globalização a imposição de uma condição de decréscimo suficiente.

6.2 Impacto real do processo de optimização

Embora numa avaliação real não seja conhecido o valor da função objectivo sem contaminação com ruído, tal é possível no caso do conjunto teste considerado. Assim, com o objectivo de verificar se os ganhos evidenciados com o uso das versões não monótonas da PDD são reais e não apenas fruto da contaminação com ruído, considerou-se a sequência de iteradas gerada pelos algoritmos mas avaliou-se o seu valor nas funções sem qualquer contaminação com ruído. Os perfis de dados correspondentes às sequências geradas com 10% de ruído encontram-se na Figura 6.4.

Como é possível observar, as estratégias não monótonas apresentam um desempenho real superior face às estratégias monótonas, mantendo-se a generalidade das conclusões extraídas na Secção 6.1, para um nível de ruído de 10%, considerando os três tipos de ruído. Isto permite afirmar que as conclusões tiradas a partir das figuras anteriores não

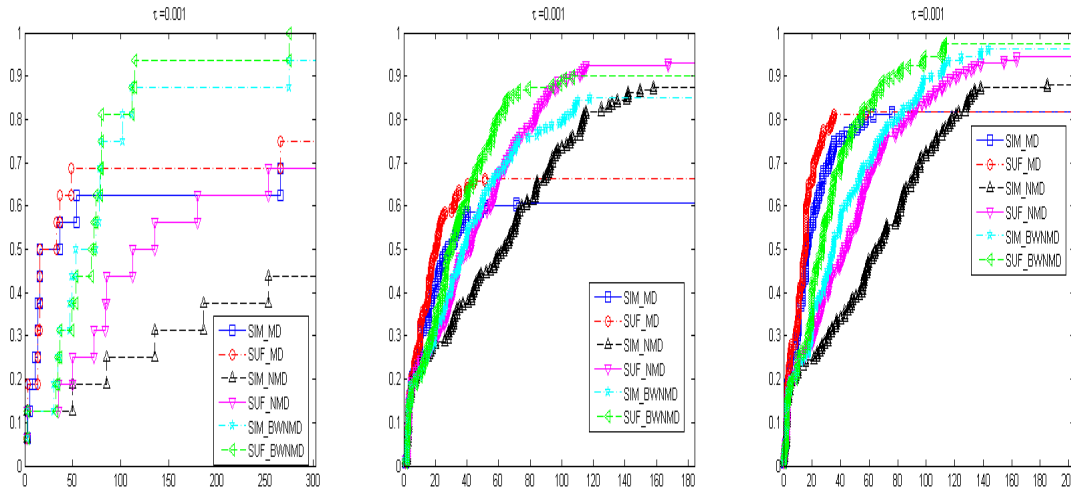


Figura 6.4: Perfis de dados para o conjunto de problemas sem contaminação com ruído, mas utilizando as sequências de iteradas geradas anteriormente pelos algoritmos, para ruído baseado em polinômios de Chebyshev, ruído Normal e ruído Uniforme, respectivamente, considerando um nível de ruído de 10%.

foram resultado dos problemas estarem contaminados com ruído, mas sim do algoritmo estar de facto a ter um bom desempenho. Note-se que poder-se-ia ter uma percepção errada acerca do bom desempenho das estratégias não monótonas porque a avaliação de pontos estava contaminada favoravelmente com ruído. Contudo, tal não sucede.

6.3 Utilização de uma ‘cache’

Como foi descrito na Secção 2.5, uma das formas de melhorar a eficiência computacional da PDD consiste na implementação de uma ‘cache’ evitando repetições na avaliação de pontos. A ‘cache’ consiste simplesmente num histórico dos pontos avaliados. Antes de proceder a uma nova avaliação, o algoritmo verifica se o ponto em causa já consta desse histórico. Caso tal se verifique, o ponto não é avaliado, evitando-se um cálculo desnecessário. Num contexto em que a avaliação da função objectivo é dispendiosa, esta simples opção pode ser crítica na melhoria da eficiência computacional. Nesta secção procura-se avaliar numericamente o efeito de uma ‘cache’, quer quando são utilizadas estratégias monótonas, quer quando se opta por estratégias não monótonas em PDD.

As opções relativas à implementação algorítmica mantêm-se, com excepção do critério de paragem. Adicionalmente ao limite máximo de 1500 cálculos, permite-se um comprimento mínimo de passo igual a $2.220E-16$. Dado que se está a considerar a implementação de uma ‘cache’, caso este critério não fosse considerado, a redução consecutiva do comprimento de passo poderia tornar impossível a distinção entre pontos, fazendo com que o limite máximo de cálculos nunca fosse atingido.

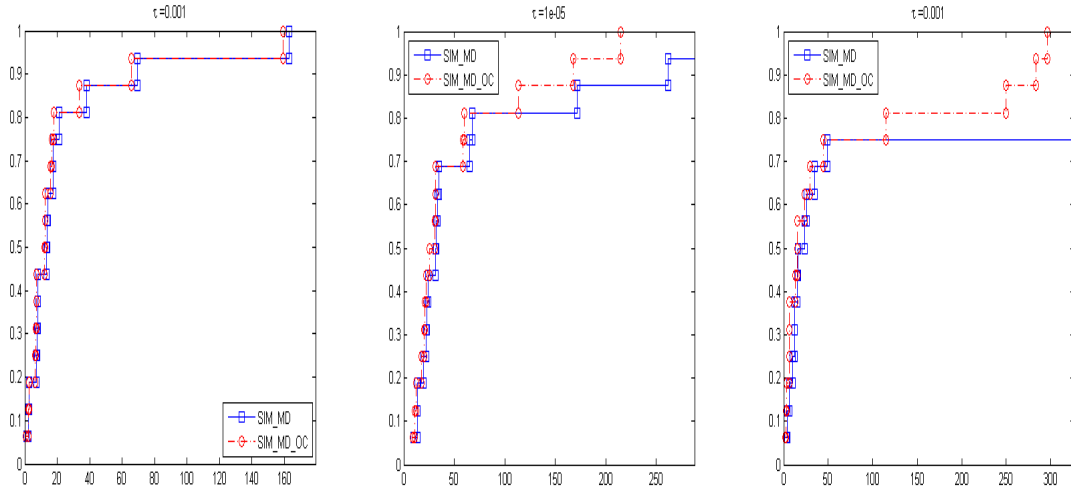


Figura 6.5: Comparação do desempenho entre a PDD utilizando uma ‘cache’ e a PDD sem ‘cache’ considerando $\beta = 2$ para os dois primeiros perfis de dados e $\beta = 1$ para o terceiro.

Na Figura 6.5 pode-se observar a comparação de desempenho entre a PDD utilizando uma ‘cache’ (*SIM_MD_OC*) e a PDD sem ‘cache’ (*SIM_MD*) na classe de problemas sem contaminação de ruído. Observa-se que, no caso de se considerar $\beta = 2$ e uma tolerância $\tau = 10^{-3}$ na elaboração dos perfis de dados, a presença de uma ‘cache’ não aparenta ter uma vantagem significativa. O mesmo não sucede se $\beta = 1$ ou se se for mais exigente no critério para considerar um problema como resolvido, fazendo $\tau = 10^{-5}$.

De facto, numa abordagem de sondagem oportunista, não aumentar o comprimento do passo em sucessos potencia a repetição de pontos, o que justifica a vantagem sentida com o uso de uma ‘cache’. Por outro lado, considerar-se $\tau = 10^{-5}$ na elaboração do perfil de dados significa uma maior exigência no critério para considerar um problema como resolvido, o que implica a necessidade de um orçamento computacional mais elevado, novamente justificando a mais valia resultante da implementação de uma ‘cache’.

Coloca-se agora a questão se a presença de ruído numérico afectará este tipo de comportamento. Para tal, repetiram-se os testes, considerando agora os três tipos de ruído (baseado em polinómios de Chebyshev, Normal e Uniforme), para um nível de 10%.

Observando a Figura 6.6 constata-se que, à excepção do ruído baseado em polinómios de Chebyshev (o único ruído determinístico), a implementação de uma ‘cache’ na PDD não parece trazer uma vantagem relevante. Uma possível explicação poderá ser o facto da PDD sem a implementação de uma ‘cache’ permitir a reavaliação de pontos, ou seja o algoritmo pode voltar a calcular pontos anteriormente avaliados o que, em problemas com ruído com uma certa aleatoriedade, pode permitir corrigir decisões, escapando a alguns mínimos espúrios resultantes do ruído. Esta pode ser a justificação para o bom desempenho que é normalmente atribuído aos algoritmos de PDD na presença de ruído [13].

Considerando um parâmetro de actualização do comprimento passo $\beta = 1$, que numa sondagem oportunista potencia a repetição de pontos, este comportamento ainda se torna

CAPÍTULO 6. ANÁLISE DE DESEMPENHO DAS ESTRATÉGIAS NÃO MONÓTONAS EM PDD NA OPTIMIZAÇÃO DE PROBLEMAS COM RUÍDO

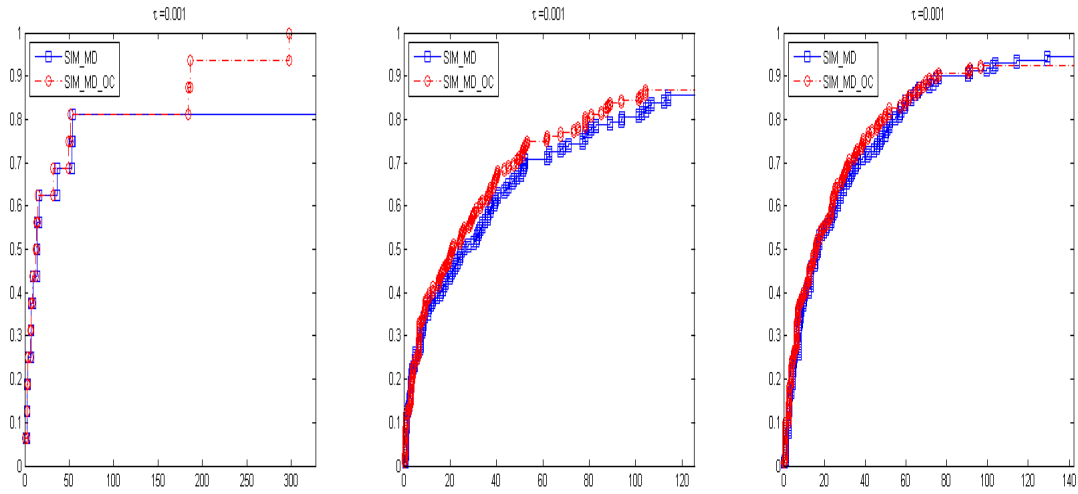


Figura 6.6: Perfis de dados para o conjunto de problemas perturbados com os três tipos de ruído (baseado em polinômios de Chebyshev, Normal e Uniforme, respectivamente), considerando um nível de ruído de 10%, com $\beta = 2$.

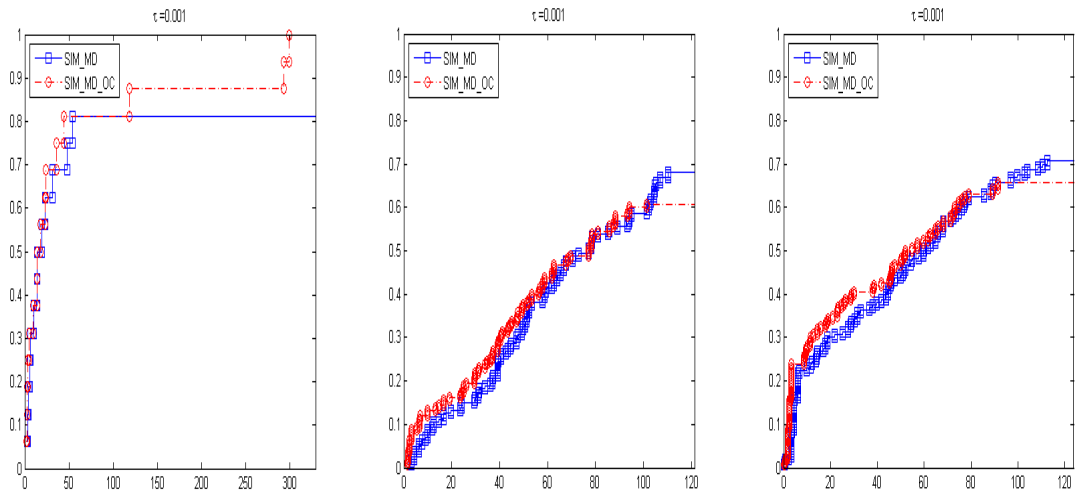


Figura 6.7: Perfis de dados para o conjunto de problemas perturbados com os três tipos de ruído (baseado em polinômios de Chebyshev, Normal e Uniforme, respectivamente), considerando um nível de ruído de 10%, com $\beta = 1$.

mais evidente conforme se pode observar na Figura 6.7.

Ainda assim, procurou-se observar qual o impacto da implementação de uma ‘cache’ em conjunto com as estratégias não monótonas. Para tal, consideraram-se os três tipos de ruído para um nível de 10%, com os comprimentos dos históricos determinados na Secção 6.1, implementando uma ‘cache’.

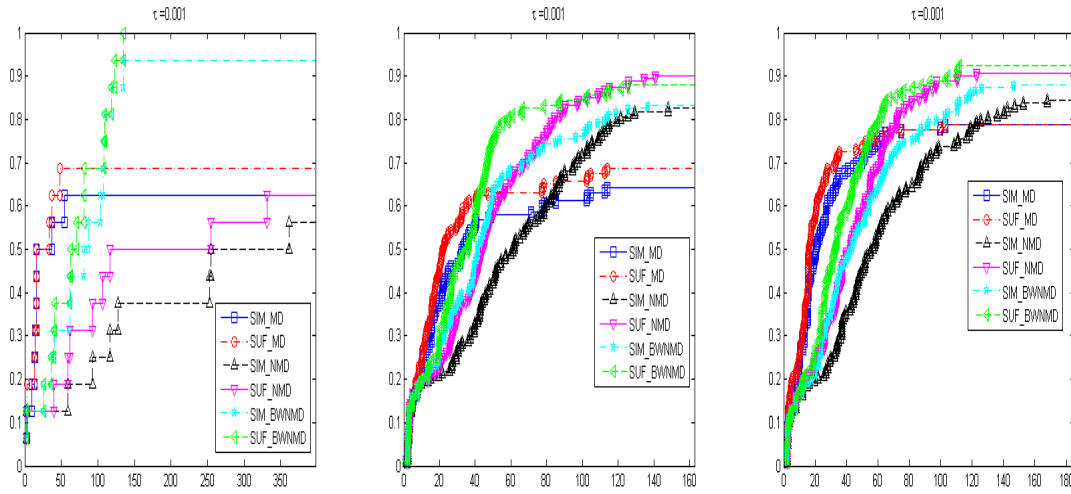


Figura 6.8: Perfis de dados para o conjunto de problemas perturbados com os três tipos de ruído (baseado em polinômios de Chebyshev, Normal e Uniforme, respectivamente), considerando um nível de ruído de 10% e a implementação de uma 'cache'.

Na Figura 6.8 pode-se observar que não existe uma mais valia significativa no uso de uma 'cache' em conjunto com estratégias não monótonas. Exceptua-se o caso do ruído baseado em polinômios de Chebyshev onde, tal como na Figura 6.6, a presença de uma 'cache' permite um desempenho ligeiramente superior.

CONCLUSÕES E QUESTÕES EM ABERTO

7.1 Conclusões

A principal contribuição do trabalho de investigação, que conduziu à escrita desta dissertação, consistiu na análise de desempenho da PDD na optimização de funções sujeitas a ruído numérico, equipando-a com uma abordagem não monótona.

Numa primeira fase, começou-se por estabelecer a necessidade da definição de estratégias de adaptação da PDD à optimização de funções sujeitas a ruído. Para tal, no Capítulo 3 procurou-se ganhar sensibilidade ao impacto que diferentes tipos de ruído (Normal, Uniforme e baseado em polinómios de Chebyshev) têm na função objectivo e à capacidade da PDD em determinar a solução dos problemas de optimização resultantes da contaminação com estes três tipos de ruído, considerando como função base uma quadrática.

Diferentes tipos de ruído afectam de forma diferente a função objectivo. O ruído baseado em polinómios de Chebyshev provoca oscilações mais regulares do que os outros dois tipos, sendo o ruído Normal o que origina oscilações mais irregulares. A percentagem de ruído presente na função objectivo (5%, 10% e 20%) afecta a distância do ponto óptimo a que se pode considerar a inicialização do método, por forma a garantir que a PDD é bem sucedida. Assim, um aumento no nível de ruído presente na função objectivo faz com que a PDD necessite de um ponto inicial mais perto do ponto óptimo para que este seja determinado. Por outro lado, concluiu-se também que o tipo de ruído Normal é o que exige inicializações mais próximas do ponto óptimo para que a PDD seja bem sucedida.

Considerando inicializações longe do ponto óptimo, esta análise foi estendida a uma classe mais completa de problemas, para os três tipos e níveis de ruído anteriormente mencionados. As conclusões tiradas deste estudo numérico foram idênticas às do estudo

anterior. Um aumento do nível de ruído presente na função objectivo afecta negativamente o desempenho da PDD. Observa-se ainda que o limite ao número máximo de cálculos permitido pela função objectivo, dentro do qual a PDD é competitiva é claramente mais baixo no caso de se considerar o ruído Normal, por comparação com os outros dois tipos de ruído considerados.

Estes estudos numéricos justificaram a necessidade da definição de estratégias que permitam aumentar o desempenho da PDD quando aplicada a problemas perturbados com ruído, principalmente no caso de serem consideradas inicializações longe do ponto óptimo. O recurso a estratégias não monótonas foi a opção tomada.

No Capítulo 5 foi feita a descrição de duas variantes algorítmicas de PDD equipadas com estratégias não monótonas, nomeadamente a *NMD* e a *BWNMD*, sendo a *BWNMD* uma estratégia mais exigente na aceitação de novos pontos, podendo ser pensada como uma versão intermédia entre a variante não monótona *NMD* e uma estratégia monótona.

Neste capítulo foi ainda estabelecida a convergência destas variantes não monótonas, seguindo a abordagem descrita na Secção 2.6 para o caso monótono. Primeiramente mostrou-se que um método de PDD usando uma variante não monótona gera uma sucessão de comprimentos de passo cujo limite inferior tende para zero (ver Teoremas 5.2.1 e 5.2.2), o que permite garantir a existência de uma subsucessão refinada convergente. Em seguida provou-se a não negatividade da derivada direccionada generalizada de Clarke da função calculada num ponto limite de uma subsucessão refinada, numa direcção de sondagem que tenha sido usada um número infinito de vezes na subsucessão refinada (ver Teorema 5.2.3). Uma versão mais robusta deste resultado foi obtida ao assumir mais suavidade acerca da função objectivo (ver Teorema 5.2.4).

No capítulo seguinte estabeleceu-se a vantagem da utilização dos algoritmos não monótonos de PDD previamente propostos, por comparação com algoritmos monótonos de PDD. Em geral, à medida que o nível de ruído aumenta, o desempenho das estratégias monótonas diminui. Constatou-se ainda que apenas orçamentos computacionais muito reduzidos justificam o uso de estratégias monótonas. Para os três tipos de ruído considerados observou-se que a variante que resolve mais problemas é a *BWNMD*, com excepção do caso em que se tem presente 20% de ruído Normal, onde a estratégia que resolve mais problemas é a *SUF_NMD*. Os comprimentos de históricos não necessitam de ser muito elevados (cerca de 5 para o ruído baseado em polinómios de Chebyshev e 2 nos restantes casos).

Com o objectivo de melhorar a eficiência computacional foi ainda analisada a possibilidade da implementação de uma ‘cache’. Em problemas onde o ruído é estocástico, a implementação de uma ‘cache’ quando são consideradas estratégias não monótonas não é uma mais valia, dada a natureza aleatória do ruído. O mesmo não acontece quando é considerado ruído baseado em polinómios de Chebyshev, onde a implementação de uma ‘cache’ permite desempenhos ligeiramente superiores.

7.2 Questões em aberto

Estabelecer a convergência de novos algoritmos de optimização e o seu bom desempenho face a algoritmos pertencentes à mesma classe é um bom começo, mas não é o suficiente para os tornar atraentes para a comunidade científica e garantir o seu uso em aplicações reais. É necessário mostrar a sua competitividade computacional face a algoritmos já existentes.

Por isso uma questão em aberto é verificar se as melhores variantes dos algoritmos propostos nesta dissertação são competitivas face a outros algoritmos de optimização sem recurso a derivadas destinados à optimização de funções sujeitas a ruído numérico (como, por exemplo, os descritos na Secção 1.2).

Dado que para orçamentos computacionais reduzidos as estratégias monótonas mostraram-se mais vantajosas, seria interessante determinar um limiar para a utilização das estratégias não monótonas, permitindo o desenvolvimento de uma variante híbrida de PDD no que respeita à condição de aceitação de novos pontos.

Seria também interessante observar como é que as variantes algorítmicas se comportam perante uma aplicação real. Para esta dissertação tentou-se contactar alguns autores de trabalhos de optimização em aplicações reais sujeitas a ruído com vista a testar os algoritmos propostos neste trabalho nas referidas aplicações. Questões de confidencialidade inviabilizaram o processo. Esta é uma questão em aberto que se gostaria de continuar a explorar.

A estrutura algorítmica inerente à procura directa direcciona e o carácter dispendioso associado à avaliação da função objectivo sugerem que reduções significativas do tempo de execução poderão ser conseguidas se se usarem múltiplos processadores. Outra questão em aberto é então a paralelização dos algoritmos propostos nesta dissertação.

BIBLIOGRAFIA

- [1] M. A. Abramson, C. Audet, J. E. Dennis, Jr. e S. Le Digabel. “OrthoMADS: A deterministic MADS instance with orthogonal directions”. Em: *SIAM J. Optim.* 20 (2009), pp. 948–966.
- [2] E. J. Anderson e M. C. Ferris. “A direct search algorithm for optimization with noisy function evaluations”. Em: *SIAM J. Optim.* 11 (2001), pp. 837–857.
- [3] L. Armijo. “Minimization of functions having Lipschitz-continuous first partial derivatives”. Em: *Pacific J. Math* 16 (1966), pp. 1–3.
- [4] C. Audet e J. E. Dennis Jr. . “Analysis of generalized pattern searches”. Em: *SIAM J. Optim.* 13 (2002), pp. 889–903.
- [5] C. Audet e J. E. Dennis Jr.. “Mesh adaptive direct search algorithms for constrained optimization”. Em: *SIAM J. Optim.* 17 (2006), pp. 188–217.
- [6] C. Audet, S. Le Digabel e C. Tribes. “Dynamic scaling in the mesh adaptive direct search algorithm for blackbox optimization”. Em: *Optim. Eng.* 17 (2016), pp. 333–358.
- [7] C. Audet, A. Ihaddadene, S. Le Digabel e C. Tribes. *Robust optimization of noisy blackbox problems using the Mesh Adaptive Direct Search algorithm*. Rel. téc. G-2016-55. Les Cahiers du GERAD, 2016.
- [8] S. C. Billups, J. Larson e P. Graf. “Derivative-Free optimization of expensive functions with computational error using weighted regression”. Em: *SIAM J. Optim.* 23 (2013), pp. 27–53.
- [9] R. G. Carter, J. M. Gablonsky, A. Patrick, C. T. Kelley e O. J. Eslinger. “Algorithms for noisy problems in gas transmission pipeline optimization”. Em: *Optim. Eng.* 2 (2001), pp. 139–157.
- [10] R. Chen e S. M. Wild. *Randomized derivative-free optimization of noisy convex functions*. Rel. téc. 1507.03332v1. arXiv, 2015.
- [11] W. Cheng e D. H. Li. “A derivative-free nonmonotone line search and its application to the spectral residual method”. Em: *IMA J. Numer. Anal.* 29 (2009), pp. 814–825.
- [12] F. H. Clarke. *Optimization and Nonsmooth Analysis*. Reissued by SIAM, Philadelphia, 1990. New York: John Wiley & Sons, 1983.

- [13] A. R. Conn, K. Scheinberg e L. N. Vicente. *Introduction to Derivative-Free Optimization*. MPS-SIAM Series on Optimization. Philadelphia: SIAM, 2009.
- [14] A. L. Custódio e L. N. Vicente. “Using sampling and simplex derivatives in pattern search methods”. Em: *SIAM J. Optim.* 18 (2007), pp. 537–555.
- [15] C. Davis. “Theory of positive linear dependence”. Em: *Amer. J. Math.* 76 (1954), pp. 733–746.
- [16] G. Deng. “Simulation-Based Optimization”. Tese de doutoramento. University of Wisconsin, Madison, 2007.
- [17] G. Deng e M. C. Ferris. “Adaptation of the UOBYQA algorithm for noisy functions”. Em: *Proceedings of the 2006 Winter Simulation Conference*. Ed. por L. F. Perrone, F. P. Weiland, J. Liu, B. G. Lawson, D. M. Nicol e R. M. Fujimoto. 2006, pp. 312–319.
- [18] G. Deng e M. C. Ferris. “Extension of the DIRECT optimization algorithm for noisy functions”. Em: *Proceedings of the 2007 Winter Simulation Conference*. Ed. por M. H. In B. Biller S. Henderson e J. Shortle. 2007, pp. 497–504.
- [19] M. A. Diniz-Ehrhardt, J. M. Martínez e M. Raydan. “A derivative-free nonmonotone line-search technique for unconstrained optimization”. Em: *J. Comput. Appl. Math.* 219 (2008), pp. 383–397.
- [20] E. Fermi e N. Metropolis. *Los Alamos unclassified report LS-1492*. Rel. téc. LS-1492. Los Alamos National Laboratory, EUA, 1952.
- [21] A. A. Giunta. “Aircraft Multidisciplinary Design Optimization Using Design of Experiments Theory and Response Surface Modeling Methods”. Tese de doutoramento. Faculty of Virginia Polytechnic Institute e State University, 1997.
- [22] S. Gratton, C. W. Royer, L. N. Vicente e Z. Zhang. “Direct search based on probabilistic descent”. Em: *SIAM J. Optim.* 25 (2015), pp. 1515–1541.
- [23] L. Grippo e F. Rinaldi. “A class of derivative-free nonmonotone optimization algorithms employing coordinate rotations and gradient approximations”. Em: *Comput. Optim. Appl.* 60 (2014), pp. 1–33.
- [24] L. Grippo, F. Lampariello e S. Lucidi. “A nonmonotone line search technique for Newton method”. Em: *SIAM J. Numer. Anal.* 23 (1986), pp. 707–716.
- [25] I. Griva, S. G. Nash e A. Sofer. *Linear and Nonlinear Optimization*. Philadelphia: SIAM, 2009.
- [26] T. Hemker. “Derivative Free Surrogate Optimization for Mixed-Integer Nonlinear Black Box Problems in Engineering”. Tese de doutoramento. Darmstadt University, Alemanha, 2008.
- [27] R. Hooke e T. A. Jeeves. ““Direct search” solution of numerical and statistical problems”. Em: *J. of the ACM* 8 (1961), pp. 212–229.

-
- [28] D. R. Jones. “Direct global optimization algorithm”. Em: *Encyclopedia of Optimization*. Ed. por C. A. Floudas e P. M. Pardalos. Boston: Springer US, 2001, pp. 431–440.
- [29] D. R. Jones, C. D. Perttunen e B. E. Stuckman. “Lipschitzian optimization without the Lipschitz constant”. Em: *J. Optim. Theory Appl.* 79 (1993), pp. 157–181.
- [30] C. T. Kelley. *Implicit Filtering*. Software Environments and Tools. Philadelphia: SIAM, 2011.
- [31] B. R. Kirsch, G. W. Characklis, K. E. M. Dillard e C. T. Kelley. “More efficient optimization of long-term water supply portfolios”. Em: *Water Resources Research* 45 (2009), W03414.
- [32] T. G. Kolda, R. M. Lewis e V. Torczon. “Optimization by direct search: New perspectives on some classical and modern methods”. Em: *SIAM Rev.* 45 (2003), pp. 385–482.
- [33] N. Krejic, Z. Luzanin e F. Nikolovski. “A nonmonotone line search method for noisy minimization”. Em: *Optim. Letters* 9 (2015), pp. 1371–1391.
- [34] J. Larson e S. C. Billups. “Stochastic derivative-free optimization using a trust region framework”. Em: *Comput. Optim. Appl.* 64 (2016), pp. 619–645.
- [35] D. H. Li e M. Fukushima. “A derivative-free line search and global convergence of Broyden-like method for nonlinear equations”. Em: *Optim. Methods Softw.* 13 (2000), pp. 181–201.
- [36] J. J. Moré e S. M. Wild. “Benchmarking derivative-free optimization algorithms”. Em: *SIAM J. Optim.* 20 (2009), pp. 172–191.
- [37] J. J. Moré, B. S. Garbow e K. E. Hillstom. “Testing unconstrained optimization software”. Em: *ACM Trans. Math. Software* 7 (1981), pp. 17–41.
- [38] J. A. Nelder e R. Mead. “A simplex method for function minimization”. Em: *The Comput. J.* 7 (1965), pp. 308–313.
- [39] F. Nikolovski e I. Stojkovska. “New derivative-free nonmonotone line search methods for unconstrained minimization”. Em: *Proceedings of the Fifth International Scientific Conference — FMNS2013, Mathematics and Informatics*. Vol. 1. Blagoevgrad, Bulgaria: South-West University “Neofit Rilski”, 2013, pp. 47–53.
- [40] M. J. D. Powell. “UOBYQA: Unconstrained optimization by quadratic approximation”. Em: *Math. Program.* 92 (2002), pp. 555–582.
- [41] V. Torczon. “Multi-Directional Search: A Direct Search Algorithm for Parallel Machines”. Tese de doutoramento. Rice University Houston, EUA, 1989.
- [42] V. Torczon. “On the convergence of pattern search algorithms”. Em: *SIAM J. Optim.* 7 (1997), pp. 1–25.

- [43] V. Torczon. “On the convergence of the multidirectional search algorithm”. Em: *SIAM J. Optim.* 1 (1991), pp. 123–145.
- [44] A. Troltsch. “An Active Trust-Region Method for Bound-Constrained Nonlinear Optimization without Derivatives Applied to Noisy Aerodynamic Design Problems”. Tese de doutoramento. INP Toulouse, França, 2011.
- [45] L. N. Vicente e A. L. Custódio. “Analysis of direct searches for discontinuous functions”. Em: *Math. Program.* 133 (2012), pp. 299–325.
- [46] H. Zang e W. Hager. “A nonmonotone line search technique and its application to unconstrained optimization”. Em: *SIAM J. Optim.* 14 (2004), pp. 1043–1056.



TABELAS

A.1 Estratégias não monótonas

Nas seguintes tabelas são apresentados os melhores valores obtidos para a função objectivo, por cada estratégia, para cada problema, para cada tipo e nível de ruído.

No caso de contaminação com ruído Uniforme ou Normal, uma vez que foram utilizadas 10 sementes em cada problema, são apresentadas estatísticas descritivas da melhor estratégia para cada nível e tipo de ruído, não considerando os resultados das sementes onde houve falhas. Considerou-se que existe uma falha quando o valor final obtido pela estratégia é muito maior do que o óptimo e do que o valor obtido pela PDD quando se considera uma inicialização longe (ver a Tabela 3.8).

Em qualquer dos casos, considerou-se o conjunto gerador positivo $[e_n, -e_n, I_n, -I_n]$, em que e_n é um vector coluna de uns com dimensão n e I_n representa a matriz identidade. Considerou-se ainda um comprimento de passo inicial $\Delta_0 = 1$ e os parâmetros $\beta = 2$ e $\mu = \frac{1}{2}$ na actualização do comprimento de passo. O passo de sondagem foi efectuado de forma oportunista, tal como descrito na Secção 2.3. No caso da estratégia de globalização ser baseada na imposição de uma condição de decréscimo suficiente, considerou-se a função $\rho(t) = t^2$. Como critério de paragem, considerou-se um máximo de 1500 cálculos para a função objectivo.

A.1.1 Ruído baseado em polinômios de Chebyshev

Tabela A.1: Mínimo obtido para cada problema e para cada estratégia, considerando um nível de ruído de 5%.

Problema	SIM_MD	SUF_MD	SIM_NMD	SUF_NMD	SIM_BWNMD	SUF_BWNMD
1	3,4275E+01	3,4275E+01	1,1453E+03	4,8565E+01	3,4277E+01	3,4275E+01
2	7,9613E+00	7,9613E+00	3,3287E+01	3,3287E+01	7,9613E+00	7,9613E+00
3	9,4390E+00	9,3869E+00	8,0491E+08	3,3250E+01	9,4402E+00	1,1565E+01
4	3,8085E+00	3,8085E+00	0,0000E+00	3,8198E+00	3,8085E+00	3,8085E+00
5	1,4554E-01	1,4554E-01	7,6192E+00	7,6192E+00	1,4554E-01	1,4554E-01
6	8,3270E+04	8,3270E+04	1,1238E+06	1,1238E+06	2,3357E+05	2,3357E+05
7	4,6735E+01	4,6735E+01	4,9007E+01	4,8383E+01	4,6735E+01	4,6735E+01
8	1,6149E+01	1,2546E-02	8,3001E+01	1,6301E+01	1,6556E+01	1,6298E+01
9	9,1365E-02	6,6870E-02	2,6921E+02	1,6474E+02	1,0906E+00	5,1974E-01
10	1,1066E+01	2,7422E+00	5,0597E+03	5,0597E+03	2,3461E+02	1,0002E+02
11	8,6577E+00	8,3108E+00	7,5802E+03	7,5802E+03	5,1300E+03	2,7887E+02
12	1,1187E+05	1,1187E+05	2,6748E+10	2,6748E+10	1,2442E+05	1,2442E+05
13	2,5010E+01	2,2596E+01	2,4606E+03	2,0104E+00	3,1640E+01	2,0103E+00
14	1,8715E+14	1,8715E+14	3,6030E+12	3,6030E+12	3,9558E+00	3,9558E+00
15	3,0679E+14	3,0679E+14	2,1264E+13	2,1264E+13	2,1010E+11	2,1010E+11
16	1,5530E+11	1,5530E+11	8,6305E+13	8,6305E+13	1,2765E+12	1,2765E+12

Tabela A.2: Mínimo obtido para cada problema e para cada estratégia, considerando um nível de ruído de 10%.

Problema	SIM_MD	SUF_MD	SIM_NMD	SUF_NMD	SIM_BWNMD	SUF_BWNMD
1	3,2495E+01	3,2499E+01	1,1866E+03	3,8229E+01	3,2495E+01	3,2536E+01
2	2,5770E+10	2,5770E+10	7,7733E+01	7,7733E+01	7,5423E+00	7,5423E+00
3	1,0914E+10	1,0914E+10	6,6491E+08	9,0526E+04	1,2446E+01	1,2436E+01
4	0,0000E+00	0,0000E+00	0,0000E+00	0,0000E+00	3,6144E+00	3,6144E+00
5	1,8585E+00	1,8784E+00	7,1721E-01	2,1140E+00	1,8587E+00	1,8805E+00
6	8,1230E+04	8,1230E+04	4,2709E+05	4,2709E+05	2,1255E+05	2,1255E+05
7	4,5428E+01	4,5428E+01	4,8013E+01	4,5432E+01	1,0076E+08	4,5428E+01
8	1,6298E+01	9,7716E-03	2,1485E+02	1,5378E+01	1,5485E+01	1,5133E+01
9	8,6723E-02	1,9625E-01	9,7135E+01	9,7135E+01	1,2273E+01	4,6770E+00
10	3,4814E+00	5,0362E+00	2,1979E+03	6,5184E+02	1,4151E+02	1,2615E+02
11	7,0588E+00	8,1389E+00	7,8874E+03	7,8874E+03	5,9089E+02	4,3760E+02
12	1,8497E+14	1,8497E+14	1,9022E+14	1,9022E+14	1,8529E+14	1,8529E+14
13	2,3678E+01	1,6162E+00	1,8452E+02	2,5721E+00	2,3699E+01	1,6341E+00
14	1,7730E+14	1,7730E+14	1,0635E+12	1,0635E+12	3,7812E+00	3,7812E+00
15	2,9083E+14	2,9083E+14	2,1207E+13	2,1207E+13	5,3017E+11	5,3017E+11
16	3,2943E+11	3,2943E+11	2,9319E+12	2,9319E+12	1,2112E+12	1,2112E+12

Tabela A.3: Mínimo obtido para cada problema e para cada estratégia, considerando um nível de ruído de 20%.

Problema	SIM_MD	SUF_MD	SIM_NMD	SUF_NMD	SIM_BWNMD	SUF_BWNMD
1	2,8895E+01	2,8907E+01	1,2691E+03	6,7224E+01	2,8909E+01	2,8944E+01
2	2,3115E+10	2,3115E+10	1,2944E+06	1,2944E+06	7,4557E+00	7,4557E+00
3	9,9099E+09	9,9099E+09	8,7541E+07	1,1961E+07	2,0896E+03	1,1074E+01
4	1,0060E+09	1,0060E+09	0,0000E+00	0,0000E+00	3,2177E+00	3,2177E+00
5	1,7228E+00	4,7716E+00	8,6769E+01	8,6769E+01	1,7228E+00	4,7716E+00
6	1,1046E+05	1,1046E+05	1,1309E+06	1,1309E+06	2,9963E+05	2,9963E+05
7	4,0467E+01	4,0467E+01	2,6481E+03	3,3006E+02	4,0467E+01	4,0467E+01
8	1,6898E-01	2,3493E-02	3,9404E+01	1,3780E+01	1,6901E-01	2,3522E-02
9	7,6769E-02	5,4651E+00	1,5571E+03	1,5571E+03	4,4965E+00	1,5048E+00
10	3,6274E+00	9,9688E+00	5,7499E+03	5,7499E+03	4,5379E+00	5,7827E+00
11	5,9175E+00	5,1292E+00	8,5018E+03	8,5018E+03	7,5646E+02	3,4946E+02
12	1,7571E+14	1,7571E+14	3,8247E+11	3,8247E+11	1,7587E+14	1,7587E+14
13	2,1009E+01	2,0596E+01	2,3860E+03	2,0170E+01	2,1396E+01	2,0310E+01
14	1,5783E+14	1,5783E+14	6,4422E+12	6,4422E+12	2,1883E+00	2,1883E+00
15	2,5851E+14	2,5851E+14	1,7236E+13	1,7236E+13	1,2220E+12	1,2220E+12
16	1,5530E+11	1,5530E+11	2,4618E+13	2,4618E+13	9,9360E+11	9,9360E+11

A.1.2 Ruído Uniforme

Tabela A.4: Estatísticas descritivas da melhor estratégia (SUF_BWNMD), considerando um nível de ruído de 5%.

Problema	Min	Med	Max	Nº Falhas
1	3,4361E+01	3,4782E+01	3,5894E+01	0
2	7,9622E+00	7,9691E+00	7,9833E+00	0
3	9,3904E+00	9,4306E+00	9,4732E+00	0
4	0,0000E+00	3,0511E+00	3,8505E+00	0
5	0,0000E+00	1,5143E-01	3,9435E-01	3
6	8,4423E+04	8,4423E+04	8,4423E+04	9
7	4,6551E+01	4,7886E+01	5,4046E+01	0
8	1,4996E+01	1,5132E+01	1,5551E+01	0
9	3,0631E-01	9,8728E-01	2,2841E+00	3
10	-	-	-	10
11	-	-	-	10
12	-	-	-	10
13	1,7057E+00	1,1834E+01	2,6639E+01	1
14	-	-	-	10
15	1,6575E+12	9,3781E+13	3,0750E+14	0
16	1,1973E+09	6,0332E+10	4,3061E+11	0

APÊNDICE A. TABELAS

Tabela A.5: Estatísticas descritivas da melhor estratégia (SUF_BWNMD), considerando um nível de ruído de 10%.

Problema	Min	Med	Max	Nº Falhas
1	3,2538E+01	3,3416E+01	3,4388E+01	1
2	7,5469E+00	7,5645E+00	7,6019E+00	1
3	8,8951E+00	8,9420E+00	8,9911E+00	0
4	0,0000E+00	1,8122E+00	3,6531E+00	2
5	0,0000E+00	7,1306E-01	3,1759E+00	3
6	1,0155E+05	1,0155E+05	1,0155E+05	9
7	4,4117E+01	4,5280E+01	5,1681E+01	0
8	1,4230E+01	1,4437E+01	1,4701E+01	1
9	1,3452E-01	1,4509E+00	3,4587E+00	0
10	3,0651E+00	5,4599E+00	7,8546E+00	8
11	-	-	-	10
12	1,0134E+06	1,0134E+06	1,0134E+06	9
13	1,6339E+00	1,8656E+01	2,5357E+01	1
14	-	-	-	10
15	1,3464E+12	1,1810E+14	2,9262E+14	0
16	1,4030E+09	7,1740E+10	2,7465E+11	0

Tabela A.6: Estatísticas descritivas da melhor estratégia (SUF_BWNMD), considerando um nível de ruído de 20%.

Problema	Min	Med	Max	Nº Falhas
1	2,9660E+01	3,0757E+01	3,2080E+01	1
2	6,7083E+00	6,7683E+00	6,8834E+00	2
3	7,9113E+00	7,9480E+00	8,0406E+00	2
4	0,0000E+00	1,9932E+00	5,0139E+00	1
5	0,0000E+00	1,7610E+00	5,0716E+00	2
6	-	-	-	10
7	3,9657E+01	4,1225E+01	4,5712E+01	2
8	1,2782E+01	1,3065E+01	1,3322E+01	1
9	3,9971E-02	1,2194E+00	3,4024E+00	2
10	7,5113E+00	1,3143E+01	1,7992E+01	6
11	-	-	-	10
12	-	-	-	10
13	1,5190E+00	1,5487E+01	2,2546E+01	1
14	-	-	-	10
15	1,6369E+12	1,3085E+14	2,6160E+14	0
16	1,2529E+09	1,3947E+11	5,9542E+11	0

A.1.3 Ruído Normal

Tabela A.7: Estatísticas descritivas da melhor estratégia (SUF_BWNMD), considerando um nível de ruído de 5%.

Problema	Min	Med	Max	Nº Falhas
1	3,1346E+01	3,3212E+01	3,8059E+01	1
2	7,0215E+00	7,1776E+00	7,3983E+00	0
3	8,3351E+00	8,8051E+00	1,0693E+01	0
4	0,0000E+00	2,4308E+00	3,4661E+00	3
5	5,5726E-03	1,9867E-01	5,8876E-01	6
6	-	-	-	10
7	3,9752E+01	4,4185E+01	5,5434E+01	0
8	1,2815E+01	1,3459E+01	1,3822E+01	0
9	1,9848E-02	6,2712E-01	1,4780E+00	4
10	9,9559E+00	1,3004E+01	1,4682E+01	1
11	-	-	-	10
12	-	-	-	10
13	1,5238E+00	1,3831E+01	2,3278E+01	0
14	-	-	-	10
15	1,5168E+12	1,0772E+14	2,7157E+14	0
16	1,0358E+09	8,3271E+10	2,7118E+11	0

Tabela A.8: Estatísticas descritivas da melhor estratégia (SUF_BWNMD), considerando um nível de ruído de 10%.

Problema	Min	Med	Max	Nº Falhas
1	2,4522E+01	3,0761E+01	5,0396E+01	1
2	5,7680E+00	6,0434E+00	6,4697E+00	0
3	6,8592E+00	7,3816E+00	7,7991E+00	2
4	0,0000E+00	2,2660E+00	4,2979E+00	1
5	7,3726E-02	9,4734E-01	2,5064E+00	5
6	-	-	-	10
7	3,1884E+01	3,8166E+01	4,9103E+01	0
8	1,0274E+01	1,1017E+01	1,1903E+01	0
9	1,8987E-01	2,4594E+00	4,4008E+00	2
10	1,9271E+00	7,2699E+00	1,0856E+01	3
11	1,1167E+01	1,1167E+01	1,1167E+01	1
12	-	-	-	10
13	1,2842E+00	8,2738E+00	1,9448E+01	0
14	-	-	-	10
15	4,5416E+11	1,1586E+14	2,1962E+14	0
16	8,6525E+08	1,3027E+11	4,3039E+11	0

Tabela A.9: Estatísticas descritivas da melhor estratégia (SUF_NMD), considerando um nível de ruído de 20%.

Problema	Min	Med	Max	Nº Falhas
1	1,6484E+01	3,7773E+01	7,1530E+01	5
2	3,2182E+00	4,5803E+00	7,2064E+00	1
3	3,6818E+00	7,2150E+00	1,9093E+01	1
4	0,0000E+00	1,2400E+00	2,9429E+00	1
5	0,0000E+00	1,1808E+00	5,0460E+00	1
6	-	-	-	10
7	1,9318E+01	3,8063E+01	6,7257E+01	5
8	4,2853E+00	5,8583E+00	7,4748E+00	1
9	4,5008E-01	6,1177E+00	9,0756E+00	4
10	6,7148E+00	1,0757E+01	1,6443E+01	6
11	-	-	-	10
12	-	-	-	10
13	1,1141E+00	5,5164E+00	1,3059E+01	1
14	-	-	-	10
15	3,4958E+12	5,8086E+13	1,2687E+14	0
16	5,4848E+08	3,3274E+14	3,3245E+15	0